

**UNIVERSIDADE FEDERAL DO PARANÁ**

**ELOISE PAVÃO GUERRA SLOMPO**

**ANÁLISE FILOGENÉTICA E FUNCIONAL DAS PROTEÍNAS RBP7 E RBP40 EM  
TRIPANOSSOMATÍDEOS**

**CURITIBA  
2014**

**ELOISE PAVÃO GUERRA SLOMPO**

**ANÁLISE FILOGENÉTICA E FUNCIONAL DAS PROTEÍNAS RBP7 E  
RBP40 EM TRIPANOSSOMATÍDEOS**

Tese de Doutorado apresentada ao  
Programa de Pós-Graduação em Biologia  
Celular e Molecular, da Universidade  
Federal do Paraná, como parte das  
exigências para a obtenção do título de  
Doutor em Ciências.

Orientador: Prof. Dr. Bruno Dallagiovanna

**CURITIBA  
2014**

Universidade Federal do Paraná  
Sistema de Bibliotecas

Slompo, Eloise Pavão Guerra

Análise filogenética e funcional das proteínas RBP7 e RBP40 em tripanossomatídeos. / Eloise Pavão Guerra Slompo. – Curitiba, 2014.  
130 f.: il. ; 30cm.

Orientador: Bruno Dallagiovanna

Tese (doutorado) - Universidade Federal do Paraná, Setor de Ciências Biológicas. Programa de Pós-Graduação em Biologia Celular e Molecular.

1. Tripanossoma cruzi 2. Proteínas 3. Expressão gênica I. Título II. Dallagiovanna, Bruno III. Universidade Federal do Paraná. Setor de Ciências Biológicas. Programa de Pós-Graduação em Biologia Celular e Molecular.

CDD (20 ed.) 574 88

# PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA CELULAR E MOLECULAR

Departamento de Biologia Celular e Departamento de Fisiologia  
Setor de Ciências Biológicas - Universidade Federal do Paraná  
Instituto Carlos Chagas (ICC/FIOCRUZ)

---

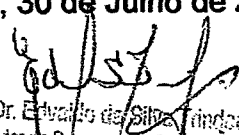
## DECLARAÇÃO

Declaramos para os devidos fins que **Eloise Pavão Guerra Slompo**, no dia 30 de Julho de 2014, no Programa de Pós-Graduação em Biologia Celular e Molecular (UFPR) defendeu sua Tese de Doutorado em Biologia Celular e Molecular, intitulada: "*Análise funcional e filogenética das proteínas RBP7 e RBP40 em tripanossomatídeos*", com a banca examinadora constituída pelos Profs: Drs. Maria Carolina Quartim Barbosa Elias-Sabbaga (Inst. Butantan), Maurílio José Soares (ICC), Wanderson Duarte da Rocha (UFPR), Alejandro Correa Dominguez e Bruno Dallagiovanna (Orientador e presidente da banca examinadora do ICC), e foi Aprovada.

A solicitação de emissão do Diploma de Doutor em Biologia Celular e Molecular ficará condicionada à implementação das correções sugeridas pelos membros da banca examinadora, bem como ao cumprimento integral das exigências estabelecidas no Regimento interno deste Programa de Pós-Graduação.

Esta declaração tem validade de 60 (sessenta) dias.

Curitiba, 30 de Julho de 2014

  
Prof. Dr. Eduardo da Silva Andrade  
Coordenador do Programa de Pós-Graduação  
em Biologia Celular e Molecular - UFPR  
Matr. 188785



Ao meu esposo Sidon Jr. e filho Aaron,  
motivação, inspiração e participação completa em  
tudo o que me proponho a fazer;

À minha família, que me depositou confiança para  
encarar todos os desafios, bem como ajuda  
constante;

Aos amigos, pelo incentivo, orações e torcida.

## AGRADECIMENTOS

A meu Deus, Criador de todas as coisas em sua mais perfeita complexidade, quem detém o universo na palma da mão, mas ainda dedica Seu incondicional amor a mim. Toda fonte de capacidade, conhecimento, inspiração e feitos. Você é o Autor, eu um instrumento.

Ao estimado orientador Bruno Dallagiovanna, meu mentor profissional, cuja opinião tenho em grande consideração, pela confiança científica neste projeto e nos paralelos, por toda dedicação ao ensino, por entender o sentido correto de exortar (estimular), pela paciência “eterna”. Responsável por parcela grande da empolgação que tenho em prosseguir cientificamente. Sou grata desde o princípio pela oportunidade de ser sua aluna, e agora parte de um grupo.

À querida colega e amiga Dra. Saloê Bispo Poubel, pela dedicação nas análises do sequenciamento, por sanar todas as dúvidas, pelas agradáveis conversas e experiência de vida.

Ao Dr. Federico Hoffmann, pelo ensino nas análises filogenéticas, pelas discussões à distância, pelo incentivo em “fazer ciência, e não crochê”.

Aos Drs. João Renato Muniz e Rinaldo Wander Montalvão, pela construção dos modelos estruturais e discussões presenciais.

À Dra. Beatriz Garat e alunos, pelo período de estágio técnico em seu laboratório, pela calorosa recepção e ensino.

Ao Dr. Samuel Goldenberg, pelos 8 anos de portas do Instituto Carlos Chagas abertas a mim, por todo empenho em construir um centro de excelência em formação, pesquisa e desenvolvimento.

Aos pesquisadores do ICC, pela constante ajuda em discussões técnicas, boa convivência, materiais confiados, crédito na minha palavra.

À equipe técnica do ICC, por prontamente preparar todos os materiais e soluções solicitados, me conferindo enorme economia de tempo.

Aos colegas do ICC, especialmente do LabREG, pela cooperação científica, discussões de trabalho, respeito e boa convivência.

*“God delights in concealing things;  
scientists delight in discovering those things.”*

Rei Salomão, versão *The Message*.

## RESUMO

Os organismos estão constantemente em modificação. Seja em resposta a pressões do ambiente, seja por eventos estocásticos, as modificações podem ser momentâneas ou hereditárias. Modificações momentâneas são desempenhadas pela regulação da expressão gênica, enquanto as hereditárias são os eventos evolutivos. A junção destas duas respostas biológicas, a evolução das redes regulatórias da expressão gênica, ainda é uma questão em aberto na biologia. O parasita unicelular *Trypanosoma cruzi*, além de sua importância médica, constitui um organismo modelo para o estudo da regulação pós-transcricional da expressão gênica. Devido à sua ramificação precoce na árvore filogenética dos seres vivos, é também um modelo no estudo da biologia evolutiva. Uma de suas RBPs (proteína de união ao RNA), a proteína regulatória TcRBP40, foi caracterizada em um estudo anterior. No atual trabalho foi determinada a sua localização celular: os reservossomos do parasita, estrutura anteriormente conhecida como de armazenamento nutricional. A presença desta proteína regulatória nesta organela, juntamente com outras proposições recentes, indica funções adicionais dos reservossomos para a biologia do parasita. Uma análise filogenética mostrou que o gene codificante da TcRBP40 originou-se junto com o gene da TcRBP7 a partir da duplicação de um gene ancestral, assim como seus homólogos em *T. brucei* (TbRBP7A e B). De acordo com a filogenia, os eventos de duplicação destes genes nestes organismos ocorreram de modo independente, após a divergência das duas espécies, e que estão acumulando mutações de modo assimétrico. As estruturas das proteínas foram preditas por modelagem por homologia e comparadas, sendo identificados os sítios divergentes. As proteínas TcRBP7 e TbRBP7A conservam maior semelhança, enquanto TcRBP40 é a mais divergente do grupo. Os alvos das proteínas TcRBP7 e TcRBP40 foram determinados por ribonômica *in vivo* e sequenciamento de nova geração, mostrando que menos de 10% são alvos em comum entre elas. Isso indica que estas RBPs de *T. cruzi*, apesar de conservarem 72% de identidade, regulam redes distintas de genes e portanto desempenham funções biológicas diferentes. Futuras análises poderão responder como as novas redes regulatórias surgem nos organismos, e como evoluem para desempenhar papéis distintos na biologia dos organismos.

## ABSTRACT

Organisms are constantly changing. In response to environmental pressures, or for stochastic reasons, those changes might be transient or inheritable. Transient changes are played by gene expression regulation, while inheritable ones consists on evolutionary events. The joining of both biological responses, which accounts for the gene expression regulatory networks evolution, is still an open question in biology. The unicellular parasite *Trypanosoma cruzi*, beyond its medical importance, is a model organism for posttranscriptional gene expression regulation studies. Due to its early-branching in the tree of life phylogeny, it is also a model for the study of evolution biology. One of its RBPs (RNA binding proteins), the regulatory protein TcRBP40, was characterized on a previous study. In this current work, its cellular location was determined: the parasite's reservosomes, previously known as a nutritional storage compartment. The presence of this regulatory protein inside this organelle, altogether with other recent propositions, indicates additional regulatory functions performed by reservosomes in the parasite's biology. A phylogenetic analysis revealed that TcRBP40 coding gene arose with TcRBP7 from the duplication of an ancestral gene, just as its *T. brucei* homologs (TbRBP7A and B). According to this phylogeny, the duplication events of these genes on those organisms occurred independently, after species divergency, and they are asymmetrically accumulating mutations. Protein structures were predicted by homology modeling and compared, identifying the divergent sites. The proteins TcRBP7 and TbRBP7A share the higher similarity, and TcRBP40 is the most divergent in this group. TcRBP7 and TcRBP40 targets were determined by *in vivo* ribonomics and next generation sequencing, showing less than 10% of overlap. This indicates that these *T. cruzi* RBPs, even though they conserve 72% identity, they regulate distinct gene networks and so play different biological roles. Other analysis may answer how the new regulatory gene networks arise in organisms, and how they evolve to play distinct roles in organisms' biology.

## SUMÁRIO

<b>RESUMO .....</b>	<b>5</b>
<b>ABSTRACT .....</b>	<b>6</b>
<b>1 INTRODUÇÃO .....</b>	<b>9</b>
<b>1.1 Evolução de Genes .....</b>	<b>10</b>
<b>1.2 Regulação da expressão gênica .....</b>	<b>13</b>
1.2.1 Relação entre evolução e regulação da expressão gênica .....	14
1.2.2 Proteínas de união ao RNA (RBPs) .....	14
1.2.3 Domínios de ligação ao RNA .....	15
1.2.4 Domínio RRM.....	18
1.2.5 Domínios de interação com o RNA e evolução .....	19
1.2.6 Motivos no RNA .....	20
1.2.7 Regulon de RNA .....	21
<b>1.3 Tripanosomatídeos .....</b>	<b>22</b>
1.3.1 Biologia .....	23
1.3.2 Ciclo de vida .....	23
1.3.3 Biologia celular .....	25
1.3.4 Biologia molecular de <i>T. cruzi</i> .....	27
1.3.5 Regulação da expressão gênica em <i>T. cruzi</i> .....	29
1.3.6 Domínio RRM em <i>T. cruzi</i> .....	31
1.3.7 TcRBP40 .....	32
<b>2 OBJETIVO .....</b>	<b>34</b>
<b>2.1 Geral .....</b>	<b>34</b>
<b>2.2 Objetivos Específicos .....</b>	<b>34</b>
<b>3 MATERIAIS E MÉTODOS .....</b>	<b>35</b>
<b>3.1 Tampões, soluções e meios .....</b>	<b>35</b>
<b>3.2 Preparações com culturas de <i>T. cruzi</i> .....</b>	<b>36</b>
3.2.1 Cultivo dos parasitas .....	36
3.2.2 Extratos proteicos de células .....	36
3.2.3 Lise solúvel de <i>T. cruzi</i> .....	37
3.2.4 Transfecção em <i>T. cruzi</i> .....	37
3.2.5 Purificação de reservossomos .....	38
3.2.6 Ensaio de endocitose .....	38
<b>3.3 Imunoensaios .....</b>	<b>39</b>
3.3.1 <i>Western blotting</i> .....	39
3.3.2 Microscopia de fluorescência.....	39
3.3.3 Imunoprecipitação.....	40
<b>3.4 Construção de vetores .....</b>	<b>41</b>
3.4.1 Amplificação dos genes .....	41
3.4.2 Vetor de entrada e vetor de expressão heteróloga.....	42
3.4.3 Vetores de fusão para transfecção em <i>T. cruzi</i> .....	43
3.4.4 Clonagem das regiões intergênicas no vetor repórter .....	44
<b>3.5 Expressão de proteínas recombinantes heterólogas .....</b>	<b>44</b>
3.5.1 Purificação de proteína recombinante solúvel .....	45
<b>3.6 Ensaio molecular de captura de RNAs por <i>pull-down</i> com proteína recombinante .....</b>	<b>46</b>
<b>3.7 Sequenciamento de nova geração .....</b>	<b>46</b>

3.7.1 Análise do sequenciamento.....	47
<b>3.8 Análises computacionais .....</b>	<b>48</b>
3.8.1 Análise de motivos na região intergênica .....	48
3.8.2 Análise filogenética.....	48
3.8.3 Predição de estrutura tridimensional.....	49
<b>4 RESULTADOS .....</b>	<b>51</b>
<b>4.1 Localização da proteína TcRBP40.....</b>	<b>51</b>
4.1.1 TcRBP40 está presente em uma fração enriquecida de reservossomos.....	51
4.1.2 A proteína TcRBP40 fusionada à GFP manteve a localização celular .....	52
4.1.3 Os sítios da TcRBP40 são organelas funcionalmente ativas .....	53
<b>4.2 O mRNA da TcRBP40 se localiza nos reservossomos .....</b>	<b>54</b>
4.2.1 O sinal de endereçamento do mRNA de TcRBP40 se encontra na região 3' UTR .....	55
4.2.2 Os mRNAs de TcRBP40 e cruzipaina compartilham motivos em potencial nas suas regiões regulatórias .....	57
<b>4.3 Análise filogenética .....</b>	<b>58</b>
4.3.1 Alinhamento da TcRBP40 contra proteínas de <i>T. brucei</i> : busca por homologia.....	58
4.3.2 Alinhamento das proteínas TbRBP7A e B contra proteínas de <i>T. cruzi</i> .....	59
4.3.3 Alinhamento das RBPs de <i>T. cruzi</i> e <i>T. brucei</i> .....	60
4.3.4 Os genes homólogos de RBP7 estão adjacentes no genoma .....	62
4.3.5 Os genes da família RBP7 em <i>T. cruzi</i> e <i>T. brucei</i> são duplicações gênicas ..	63
4.3.6 O gene de RBP7 é também duplicado em outros tripanossomatídeos.....	63
4.3.7 Os genes de TbRBP7A e TbRBP7B acumulam mutações assimetricamente ..	66
<b>4.4 Comparação molecular entre as RBPs.....</b>	<b>67</b>
4.4.1 Modelagem molecular.....	68
<b>4.4 Caracterização funcional das proteínas TcRBP7 e TcRBP40 .....</b>	<b>76</b>
4.5.1 Os genes de TcRBP7 e TcRBP40 apresentam perfis proporcionais de expressão em epimastigotas e tripomastigotas metacíclicos.....	76
4.5.2 TcRBP7 e TcRBP40 apresentam a mesma localização celular .....	77
4.5.3 Os genes de TcRBP7 e TcRBP40 são parálogos .....	79
4.5.4 Ribonômica <i>in vivo</i> .....	79
4.5.5 Ribonômica <i>in vitro</i> .....	90
4.5.6 Análise funcional dos alvos de TcRBP7 e TcRBP40 .....	91
<b>5 DISCUSSÃO.....</b>	<b>94</b>
<b>5.1 Localização da TcRBP40 e os Reservossomos .....</b>	<b>94</b>
<b>5.2 Filogenia da família RBP7 em tripanosomas.....</b>	<b>98</b>
<b>5.3 Divergência funcional de TcRBP7 e TcRBP40 .....</b>	<b>100</b>
<b>CONCLUSÃO .....</b>	<b>103</b>
<b>PERSPECTIVAS .....</b>	<b>104</b>
<b>REFERÊNCIAS .....</b>	<b>105</b>
<b>ANEXO .....</b>	<b>112</b>

## 1 INTRODUÇÃO

A visão dos sistemas biológicos à luz da presente era gênômica evocou questões antigas especuladas pela ciência. Quais seriam os determinantes da diversidade observada no mundo natural? Partindo do princípio de que todas as coisas se modificam ao longo do tempo, pode-se extrapolar que as forças que determinam as modificações se aplicam a todos os organismos, nos seus mais distintos nichos. Cada gene, cada via, cada sistema está sujeito a pressão, ou seja, interferências de um mundo físico. O fato de as formas biológicas interagirem entre si e compartilharem informações aumenta o escopo das interações físicas para um novo nível de grandeza: a comunicação biológica. Ela é responsável por interferências que não são explicadas por fenômenos ambientais, e muitas das interações biológicas estão ainda em fase de descoberta e exploração, enquanto outras permanecerão indefinidamente ocultas ao conhecimento. As descobertas científicas na biologia acompanham o desenvolvimento tecnológico das gerações, e é possível que vagas ideias do passado venham a se confirmar no futuro, quando estas grandezas discutidas vierem a ser exploradas.

Por hora, o que pode-se especular é acerca do momento científico em que nos encontramos. A ciência biológica está finalmente podendo explorar os organismos a nível de sistemas, e não mais apenas pela abordagem minimalista (KITANO, 2002). O avanço das metodologias em larga-escala tem permitido desvendar grandes segredos biológicos, ocultos até há duas décadas.

É pretendido aqui mostrar um relance do que pode vir a ser parte de uma importante descoberta científica, pois estudar as dinâmicas das modificações naturais herdadas, a saber a evolução das redes regulatórias, é algo que ainda intriga pensadores mesmo de outras áreas do conhecimento. Eis então mais uma pincelada no painel do conhecimento sobre os sistemas biológicos.



## 1.1 Evolução de Genes

A definição clássica de gene o conceitua como um segmento de DNA contendo uma região codificante capaz de gerar um produto funcional. Para a bioinformática, são regiões delimitadas por elementos sequenciais capazes de serem transcritos, ou o conceito básico de ORF (*Open Reading Frame*). Assim, necessitam ter um sítio de início e parada transcricional. Adicionalmente, regiões regulatórias à montante sinalizam sua presença e determinam a ativação/repressão da transcrição (ALBERTS et al., 2010).

A função do gene para a célula é meramente informativa. O produto de sua leitura será uma mensagem codificada ou uma molécula funcional. Assim, o sucesso de um gene não depende da sua própria natureza, mas do desempenho da função do seu produto, o qual, para ser gerado, depende do produto de outros genes.

Eventos de mutação gênica são modificações ocorridas no DNA, seja pela inserção, deleção ou troca de nucleotídeos isoladamente ou de um segmento. Estes eventos ocorrem de modo aleatório por erros de replicação, transposição de elementos ou interferência ambiental. O grande determinante da permanência de uma forma mutada de um gene se dá pelo efeito que ela exercerá sobre o seu produto final, ou seja, o quanto irá afetar a função desempenhada por este produto. Genes codificantes de proteína podem sofrer mutações redundantes, ou sinônimas, quando não há troca de aminoácido na sequência peptídica devido à redundância do código genético. Mutações não-sinônimas são aquelas que provocam a troca do aminoácido em determinada posição, podendo afetar sítios de interação ou a estrutura do peptídeo e comprometendo sua função original. Mutações sinônimas podem também ser críticas a nível de transcrito, podendo alterar a estrutura deste ou sítios de reconhecimento por outros elementos, responsáveis pela destinação do mensageiro.

O fenômeno do surgimento de novos genes pode ser explicado por dois mecanismos distintos. O primeiro deles é o surgimento *de novo*, ou seja, a criação de elementos estruturais de proteínas por mutações de regiões não-codificantes. Até há poucos anos pensava-se que este fenômeno seria excepcional na natureza (LONG et al., 2003). No entanto, recentemente foi observado entre espécies próximas de *Drosophila* que genes não-codificantes passaram a ser expressos, conferindo alguma

vantagem reprodutiva para que sejam estabelecidos na população. Os autores sugerem que estes novos genes são formados quando uma mutação aleatória na maquinaria de regulação faz com que regiões do DNA que não seriam codificadas sejam transcritas em RNA (ZHAO et al., 2014). No entanto, não se sabe o quanto deste fenômeno é realmente representativo na geração de novos genes.

O segundo mecanismo de surgimento de novos genes codificantes é a duplicação gênica, que consiste no aumento do número de cópias de um segmento de DNA. A extensão genômica desta duplicação pode ser muito variável, incluindo desde uma duplicação interna de um gene, duplicação de um gene completo, de um locus cromossômico, cromossomo completo ou até de um genoma completo, dando origem à poliploidia (PATTHY, 2007).

Um evento de duplicação gênica é frequentemente seguido de evolução acelerada da sequência e rearranjo do gene, o que oculta a detecção da fonte original. Ainda assim, pôde ser demonstrado que eventos de duplicação gênica não são uniformes em um organismo. As transições de organização dos genes são acompanhadas por um irrompimento de duplicações diante de um enfraquecimento da seleção purificadora (KOONIN, 2009).

A permanência de uma duplicação em uma espécie é determinada pela consequência de se ter uma cópia extra do gene ou região duplicados, podendo acarretar em um aumento na quantidade do produto desse gene/região. As duplicações, como as mutações, podem ser eventos funcionalmente vantajosos, neutros ou deletérios. Sua permanência na espécie depende da pressão seletiva exercida sobre os genes e do efeito causado pelo aumento nos níveis do seu produto (PATTHY, 2007).

O significado evolutivo da duplicação gênica é dar origem a uma réplica redundante de um gene, que pode divergir por mutações e eventualmente originar um novo gene, a neofuncionalização. A duplicação gênica ocupa assim uma importante função no surgimento de novos genes e novas funções. Esta assertiva é sustentada pelo fato de a maioria dos genes nos genomas possuir parálogos, indicativos de duplicação em algum ponto do processo evolutivo (KOONIN, 2009).

No entanto, o destino das cópias de um gene não é sempre o sucesso. Aliás, a probabilidade de sucesso é realmente baixa. Sendo redundantes, estão grandemente

sujeitos a mutações que causam perda da função original do produto, mas não eliminação do gene na população, fenômeno chamado de pseudogenização. Após algum tempo, os pseudogenes são deletados da população ou são tão diferentes do seu original que não são mais identificáveis (ZHANG, 2003). No entanto, alguns pseudogenes podem “renascer” e tornar a codificar o produto, ou ainda ser transcritos e ter como produtos elementos regulatórios, o que questiona sua definição original (BALAKIREV; AYALA, 2003).

Quando há certa vantagem evolutiva na duplicação, as cópias podem ser mantidas intactas pela pressão da seleção purificadora. Isto ocorre em genes cujos produtos são necessários em grande abundância na célula. De outro modo, se o produto não é abundante, é improvável que as cópias serão mantidas idênticas. Neste caso pode ocorrer a subfuncionalização, quando as cópias divergem, por exemplo, na expressão gênica, atuando em células ou compartimentos diferentes, ou na qualidade da função, tendo melhor desempenho o produto de uma delas (NOWAK et al., 1997).

Há duas opiniões distintas quando se trata do surgimento de novas funções para genes duplicados. Uma primeira ideia é de que a duplicação gênica é seguida de um relaxamento da pressão seletiva sobre um gene, o que facilita sua evolução e acelera o ganho de nova função. O contra-argumento é de que ambas as cópias sofrem esse relaxamento de restrição evolutiva de modo mais ou menos equivalente, e isso levaria a uma evolução desmedida e perda da função original por parte de ambas as cópias. Diferentemente, a outra linha sustenta que o caminho mais comum após uma duplicação seja a subfuncionalização, e não a neofuncionalização. Há sim um relaxamento de restrição evolutiva, e os parálogos retêm parte da função original em cada um, enquanto o restante da função se deteriora com o acúmulo de mutações (LYNCH; FORCE, 2000; LYNCH; KATJU, 2004). Porém, análises mais sofisticadas mostram que ambos os regimes podem acontecer nos genes parálogos. Primeiramente ocorreria uma rápida subfuncionalização, para então uma lenta neofuncionalização (CONANT; WOLFE, 2008; HE; ZHANG, 2005). Então a rota para o surgimento de novos genes seria de que os genes duplicados tendem a rapidamente diversificar os novos parálogos. Desta maneira, é como se gerasse as novas funcionalidades ajustando as antigas, mas mantendo uma cópia de segurança (KOONIN, 2009).

## 1.2 Regulação da expressão gênica

A percepção do ambiente é uma característica inerente à vida celular. Traçando um paralelo com a evolução, a regulação da expressão gênica pode ser considerada uma resposta adaptativa ao ambiente ainda em vida. Os seres vivos precisam apresentar respostas imediatas frente aos ambientes que ocupam. Enquanto nos organismos multicelulares, como o ser humano, a origem e localização de cada célula é que determina suas funções desempenhadas, nos unicelulares o ambiente é o determinante do seu comportamento e trabalho. Um parasita obrigatório depende de um ambiente específico e restrito para sua sobrevivência e propagação. Um organismo de vida livre pode encontrar diferentes condições e deve estar preparado para se adequar a cada uma no objetivo de sobreviver e se propagar, ou preparar-se para escapar de uma condição adversa quando a detectar.

As proteínas, as quais atuam nos diferentes processos metabólicos, estruturais e regulatórios em um organismo, são produzidas a partir da codificação da mensagem contida nos genes. Estes genes não atuam sozinhos: dependem de sinais intrínsecos e/ou externos para sua ativação/repressão e para determinar a taxa ou extensão desta ativação/repressão. Os sinais externos à molécula de DNA são os reguladores da expressão gênica. Os reguladores podem atuar tanto no nível transcricional, que é o início do processo de encaminhamento da mensagem, como pós-transcricional, após a mensagem ter sido copiada do DNA e levada para seu destino final: a tradução de uma mensagem nucleotídica para uma sequência peptídica.

A regulação da expressão gênica ganhou um maior respeito na era pós-genômica. A principal percepção obtida dos projetos do início do século foi de que conhecer quais são os genes e suas possíveis variações em uma espécie ou condição não trouxe a explicação para a complexidade apresentada pelos organismos. Hoje há um grande esforço científico em conhecer e entender todos os mecanismos que atuam na determinação dos níveis de produção das proteínas, além do nível transcricional.

O estudo da regulação pós-transcricional da expressão gênica vem sendo bombardeado por novas descobertas: da função de RNAs dos quais sequer se conhecia a existência (CECH; STEITZ, 2014); da bidirecionalidade do dogma central da biologia (KOONIN, 2012; SABIN; DELÁS; HANNON, 2013); de regiões não

codificantes no DNA, anteriormente rebaixadas ao título de “lixo”, que se vêm transcritas e influenciando condições de expressão (MATTICK, 2009)) de proteínas que atuavam como enzimas em uma determinada via, mas se ligam a mensageiros e promovem alteração na sua meia-vida (SCHERRER et al., 2010); de vias inteiras de regulação pós-transcricional até há pouco desconhecidas (FIRE et al., 1998; MONTGOMERY; XU; FIRE, 1998). Há um grande terreno a ser explorado mesmo em organismos já tão bem caracterizados, sendo agora vistos sob uma nova perspectiva molecular.

### *1.2.1 Relação entre evolução e regulação da expressão gênica*

Através de novos dados de biologia de sistemas foi possível observar que a expressão gênica exerce um *feedback* substancial na evolução gênica. A hipótese do mal-dobramento induzido por mal-tradução (MIM) considera que a causa dominante da taxa de evolução de uma sequência é determinada pela taxa de eventos traducionais: o nível de expressão gênica. Desta forma, a taxa de tradução pode ser vista como um amplificador do custo adaptativo do ajuste de uma proteína (KOONIN, 2009; PATTHY, 2007).

Há uma anti-correlação entre a taxa de expressão gênica e a taxa de evolução de uma proteína. Observa-se que as proteínas altamente expressas evoluem lentamente, e isto não é devido à importância da sua função para a célula. Apesar de várias hipóteses trabalharem no tema, nenhuma apresentou dados suficientes para explicar a maioria dos casos de anti-correlação. Um trabalho recente mostrou que o dobramento dos mRNAs abundantes está sob maior força de seleção, e tem impacto diferencial nas taxas de substituição sinônimas e não-sinônimas (PARK et al., 2013).

Tanto as taxas de expressão de um mRNA quanto o seu dobramento estão relacionados aos elementos em *trans* da regulação da expressão gênica: as proteínas que os reconhecem.

### *1.2.2 Proteínas de união ao RNA (RBPs)*

RBPs são proteínas que se ligam ao RNA, com afinidade por elementos de sequência ou ao esqueleto da molécula, e são responsáveis por promover

processamento, destinação, ativação ou degradação dos transcritos, sejam eles codificantes ou não. Desta forma, no nível pós-transcricional da expressão gênica, as RBPs são um dos principais agentes executores do controle de toda a existência do RNA (AUWETER; OBERSTRASS; ALLAIN, 2006; SANCHEZ-DIAZ; PENALVA, 2006).

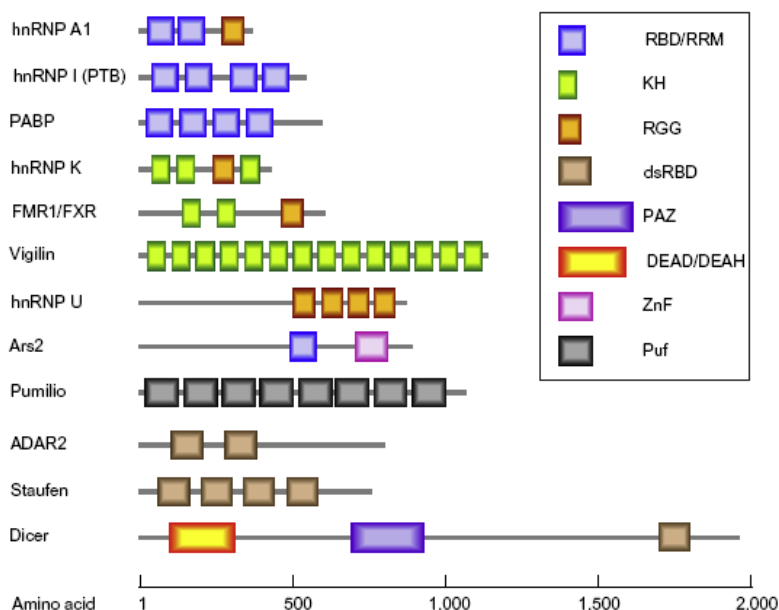
Uma das evidências da importância das RBPs para os organismos eucariotos é o fato de que uma considerável parte do seu genoma codifica RBPs. Leveduras contêm de 5 a 8% de seus genes anotados para RBPs, e em *Drosophila* chegam a quase 2% (GLISOVIC et al., 2008). Isto sugere a importância que estas proteínas podem ter na regulação pós-transcricional da expressão gênica.

A presença de domínios acessórios ou regiões especializadas nas RBPs lhes permite interagir com outras proteínas e outras moléculas. Desta forma, podem ser ativadas ou inibidas por um determinado ligante, ou ser uma ponte entre o reconhecimento específico de um mensageiro e uma proteína efetora, cuja função pode ser de levar à tradução, transporte ou degradação do transcrito. O conjunto formado pelas RBPs, mais os RNAs com os quais estão interagindo e as proteínas acessórias associadas às RBPs é chamado complexo ribonucleoproteína (RNP).

Os RNPs são montados de acordo com as proteínas e mensageiros disponíveis na célula no momento. Assim, uma alteração ambiental pode desencadear a reconfiguração dos RNPs, de acordo com o estímulo detectado, podendo eles inclusive exercer uma função oposta por conta do rearranjo, como estabilidade x degradação. Assim, os complexos RNP são responsáveis pela movimentação regulada dos RNAs nas células (KEENE, 2001).

### 1.2.3 Domínios de ligação ao RNA

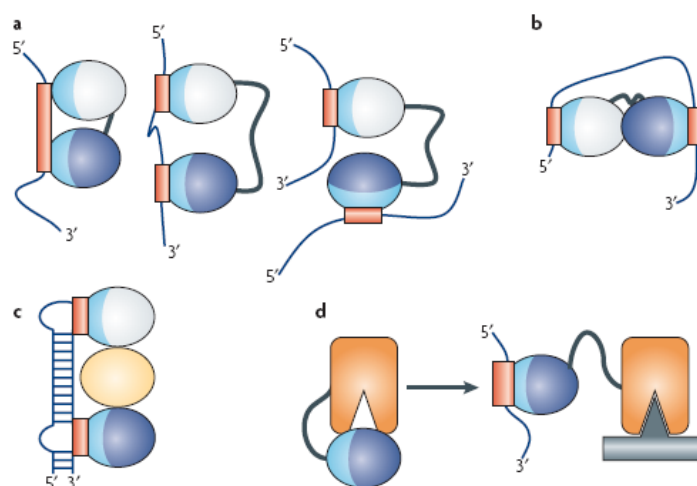
Para abranger a diversidade de funções das RBPs se faz necessária uma diversidade de estruturas de reconhecimento do RNA. As RBPs conhecidas possuem vários domínios caracterizados como sendo de interação com ácidos ribonucleicos (Fig. 1). Já foram descritos na literatura mais de 40 domínios de ligação ao RNA (AUWETER; OBERSTRASS; ALLAIN, 2006).



**Figura 1.** Domínios de ligação ao RNA nas RBPs. Frequentemente, vários domínios de ligação ao RNA são encontrados em uma RBP. Diferentes domínios incluem o *RNA Recognition Motif* (RRM), *K-homology domain* (KH), *Arg-Gly-Gly box* (RGG), *double stranded RNA-binding domain* (dsRBD), *Piwi/Argonaute/Zwille domain* (PAZ), *RNA helicase DEAD/DEAH box*, *RNA-binding zinc-finger* (ZnF) e *Puf RNA-binding repeats* (Puf). Fonte: (GLISOVIC et al., 2008).

O domínio *K homology* (KH), inicialmente visto na partícula hnRNP K, pode se associar a moléculas de fita simples de DNA ou RNA. O *Zinc finger* é também um modelo clássico de ligação a ácidos nucleicos. A subfamília do motivo CCCH é especializada para ligação em RNA. Já o domínio *Double-stranded RBD* (dsRBD) interage com RNAs de dupla-fita pelo esqueleto externo da molécula, sem especificidade por bases. O domínio *Pumílio* (PUF) já teve seu código de interação decifrado, sabendo-se quais aminoácidos são capazes de interagir com determinadas sequências de RNA (DONG et al., 2011). As proteínas da família PUF se associam à região 3'UTR de transcritos e influenciam sua estabilidade e tradução. O domínio S1 foi primeiramente identificado na subunidade ribossomal, posteriormente encontrado em outras RBPs e mesmo exonucleases. E o domínio *Piwi/Argonaute/Zwille* (PAZ) está presente em proteínas envolvidas no processamento de miRNAs, posicionando a molécula para clivagem pela Dicer (Fig. 1) (AUWETER; OBERSTRASS; ALLAIN, 2006; LUNDE; MOORE; VARANI, 2007).

Os domínios de interação com o RNA são particulares em estrutura e combinações possíveis, podendo atuar isoladamente ou em conjunto, criando superfícies versáteis para definir especificidade, afinidade e estabilidade na interação com o alvo (Fig. 2) (LUNDE; MOORE; VARANI, 2007). As forças físicas envolvidas nesta interação entre os domínios proteicos e os RNAs são do tipo fracas, envolvendo as interações eletrostáticas, responsáveis pela atração inicial que reúne RNA e proteína, enquanto empilhamento e pontes de hidrogênio mantêm o RNA em sua orientação adequada no complexo RNA-proteína (AUWETER; OBERSTRASS; ALLAIN, 2006).



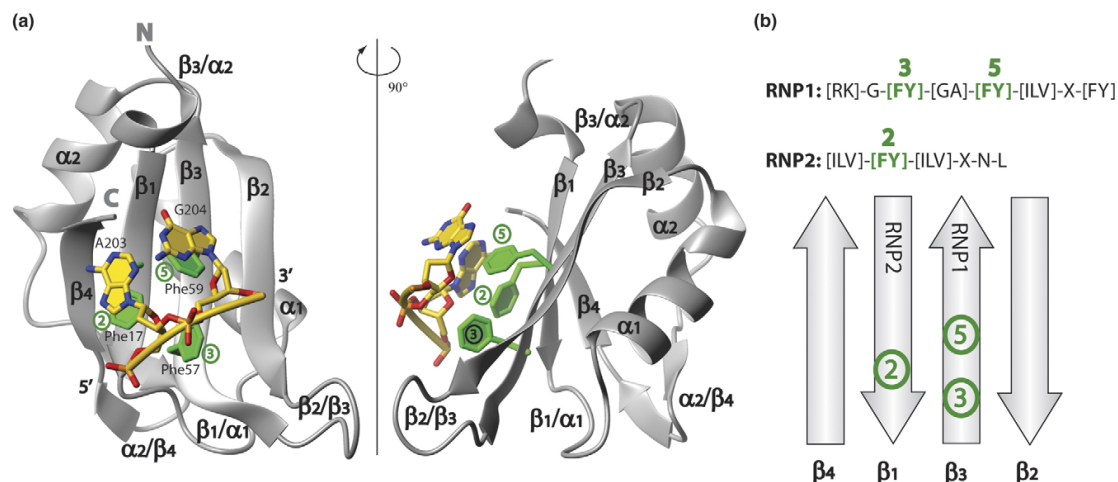
**Figura 2.** Módulos de ligação ao RNA são combinados para criar múltiplas aplicações funcionais. Domínios de ligação ao RNA funcionam de várias maneiras. (a) Reconhecem sequências de RNA com afinidade e especificidade tais que não seria possível para um único domínio ou se múltiplos domínios não cooperassem. Reconhecem uma longa sequência de RNA (esquerda), sequências separadas por vários nucleotídeos (centro), ou sequências que pertencem a moléculas distintas (direita). (b) Os domínios podem organizar mRNAs topologicamente interagindo simultaneamente com múltiplas sequências de RNA. (c) Alternativamente, podem atuar como espaçadores para posicionar corretamente outros módulos para reconhecimento. (d) Podem interagir com domínios enzimáticos para definir a especificidade do substrato para a catálise ou para regular a atividade enzimática. Os módulos de ligação ao RNA são representados por elipses com sua superfície de contato com o RNA marcada em azul-claro, e o sítio correspondente no RNA marcado em vermelho; domínios individuais estão marcados diferentemente. Fonte: (LUNDE; MOORE; VARANI, 2007).



### 1.2.4 Domínio RRM

Dentre os domínios já caracterizados, o mais comumente encontrado nas RBPs e também o mais estudado é o *RNA Recognition Motif* (RRM, Motivo de Reconhecimento do RNA). Encontrado em mais de 50% das RBPs, uma proteína que contenha este domínio pode ser automaticamente considerada uma RBP. Sendo as mais abundantes, as proteínas da família RRM estão envolvidas na maior parte dos processos de regulação pós-transcricional, senão em todos (LUNDE; MOORE; VARANI, 2007). Não obstante, o RRM é mais conhecido pela sua presença em proteínas envolvidas com *splicing*, como a PTB e a U1A, pois foram estas as primeiras proteínas descritas da família (DEKOSTER; DELANEY; HALL, 2014; KAFASLA et al., 2012).

O domínio RRM é composto por aproximadamente 90 aminoácidos, arranjados em 4 fitas- $\beta$  antiparalelas entremeadas por 2  $\alpha$ -hélices ( $\beta\alpha\beta\alpha\beta$ ) (Fig. 3a). Nesta estrutura há 2 motivos conservados de interação com o RNA: RNP1 e RNP2 (Fig. 3b). A interação com a molécula de RNA se dá através de 3 resíduos conservados nas fitas- $\beta$ , localizados nestes motivos conservados: uma Arg ou Lys que forma uma “ponte de sal” com o esqueleto fosfodiéster; e 2 resíduos aromáticos, que fazem interações por empilhamento das nucleobases. Outras forças podem atuar na interação RNA-proteína, como interações eletrostáticas e hidrofóbicas. De 2 a 8 nucleotídeos podem ser reconhecidos, dependendo da estrutura apresentada pelo RNA (AWETER; OBERSTRASS; ALLAIN, 2006;) No entanto, nos RRM de proteínas humanas, outras topologias para o reconhecimento de RNAs usando outras regiões do domínio já foram reportadas, inclusive regiões de *loops*, que são regiões que conectam as estruturas secundárias (CLÉRY; BLATTER; ALLAIN, 2008; DAUBNER; CLÉRY; ALLAIN, 2013).



**Figura 5.** Estrutura tridimensional do domínio RRM e sua interação com ácido nucleico. a) Modelo da interação tridimensional do domínio RRM2 da proteína hnRNP A1 (em cinza), interagindo com DNA telomérico simples-fita (em cores). b) Esquema da folha- $\beta$  e posição dos seus motivos conservados de interação com o ácido nucleico: RNP1 e RNP2. Os números destacam a posição dos aminoácidos nestes motivos conservados que estão interagindo diretamente com a molécula de DNA nesta estrutura. Fonte: (CLÉRY; BLATTER; ALLAIN, 2008).

### 1.2.5 Domínios de interação com o RNA e evolução

Os domínios de interação com o RNA, em geral, são conservados evolutivamente entre os organismos. A maioria deles requer uma estrutura tridimensional muito específica, como é o caso dos domínios dedo de zinco, estrutura mantida pela interação eletrostática entre cisteínas e histidinas em  $\alpha$ -hélice e fitas  $\beta$ , e um átomo de zinco. A variação de um único aminoácido dentro deste conjunto pode alterar a afinidade do domínio por moléculas de DNA para RNA (CHEN; VARANI, 2013).

Outro exemplo são as proteínas Sm, proteínas de união ao RNA que atuam na via de degradação dos mRNAs. São altamente conservadas entre taxa eucarióticos distantes. O modelo inicial da origem da família nestes organismos propôs que derivaram dos procariotos, seguindo o processo de diversificação-duplicação. Como sua função está intimamente relacionada à estrutura do RNA, a evolução das subunidades que compõem o anel das Sm é restringida pela interação transiente que deve ter com a molécula. A interação com o snRNA é dependente da heterogeneidade do anel, formado por diferentes proteínas Sm. Isso deu certa liberdade à

diversificação das Sm, que puderam desenvolver cadeias laterais funcionalmente diferentes (SCOFIELD; LYNCH, 2008).

O domínio RRM é considerado conservado estruturalmente, mas apresenta maior variação aminoacídica em relação às proteínas Sm. Isto pode ser explicado pela natureza diferente da interação do domínio com o RNA, que é independente da estrutura mas tem especificidade por elementos no RNA que podem ser variáveis, sejam eles de sequência ou estruturais.

Por esta razão, pode-se esperar maior variabilidade dentro da família RRM, podendo inferir dentro de um organismo quais RBPs contendo RRM tiveram origem comum, ou seja, os genes parálogos codificantes de domínios RRM no organismo, apesar da conservação estrutural do domínio.

#### 1.2.6 Motivos no RNA

Os RNAs são polímeros compostos por moléculas de ácidos ribonucleicos, que podem ter 4 bases nitrogenadas distintas, cuja sequência é determinada pelo DNA molde. São moléculas flexíveis, formando interações secundárias entre regiões de sua sequência.

O reconhecimento da mensagem contida nos RNAs é realizado por moléculas que se associam a elementos nas regiões codificantes ou nas regiões não traduzidas (UTRs – *UnTranslated Regions*). Estes elementos de reconhecimento nos RNAs são chamados elementos em *cis*, que funcionam como um código para determinar as funções e a vida dos RNAs. Podem ser sequenciais, determinados por uma combinação de bases, ou estruturais, formados a partir de interações secundárias e até terciárias entre regiões da própria molécula. Os motivos presentes na sequência do RNA reconhecidos por proteínas podem estar localizados em regiões não-codificantes 3' e 5' ou em regiões codificantes, apesar destas serem menos comuns. Estes elementos são reconhecidos pelos vários domínios de ligação ao RNA. Alguns exemplos são o motivo YCAY, reconhecido pela proteína neuronal humana Nova, o motivo UAGG, de ligação à hnRNP A, e o UCUU para a PTB. Em células de mamíferos, vários mRNAs instáveis contêm elementos ricos em AU (AREs – *A-U rich elements*) na sua região 3' não-codificante (MIGNONE et al., 2002). Uma

proteína humana que reconhece este motivo é a ELAV, que é expressa predominantemente em neurônios e está associada a doenças (KEENE, 2001).

### *1.2.7 Regulon de RNA*

Os RNAs são regulados por proteínas que a eles se associam. As RBPs, por sua vez, são reguladas por outras proteínas e moléculas, tanto em atividade quanto em abundância. Esta associação é dependente de respostas que a célula exibe ao ambiente, e em relação ao momento em que se encontram no ciclo de vida. Isso faz das RBPs importantes reguladores das funções celulares, uma vez que dependem de proteínas em taxas e localizações precisas.

A regulação pós-transcricional é entendida como sendo uma resposta mais acelerada a um estímulo, pois atua diretamente nos transcritos, eliminando o “atraso” da síntese de mRNAs quando uma rápida resposta é solicitada. Assim, os mensageiros cujos produtos devem atuar na mesma via metabólica requerida devem conjuntamente serem traduzidos, enquanto produtos participantes de vias inativas são regulados negativamente.

De acordo com a teoria de operons pós-transcricionais eucarióticos, mRNAs funcionalmente relacionados, ou seja, codificantes de proteínas pertencentes a uma mesma via metabólica, são regulados combinatoriamente por um conjunto de RBPs (KEENE; TENENBAUM, 2002). Assim, estes mensageiros são encontrados em um mesmo complexo RNP, formado por proteínas de união ao RNA, regulatórias, e proteínas acessórias da maquinaria de regulação. A teoria já foi comprovada em diversos organismos, inclusive em tripanossomatídeos, organismos da base da árvore filogenética (NOÉ; GAUDENZI; FRASCH, 2008). Esses complexos RNPs contendo RNAs relacionados funcionalmente podem se rearranjar de acordo com fatores que interfiram na regulação da expressão gênica. A esta propriedade combinatorial entre os complexos RNPs foi dado o nome regulon de RNA (KEENE, 2007).

Para conhecer os diferentes complexos RNP de um organismo e poder mapear seu regulon é necessário identificar as RBPs existentes e seus RNAs alvos. A identificação de RBPs se dá através de metodologias de bioinformática, pela busca por domínios conservados de interação com o RNA (GAUDENZI; FRASCH;

CLAYTON, 2005), ou por metodologias funcionais já descritas (BUTTER et al., 2009). A identificação dos alvos pode se valer de vários métodos de captura seguidos de sequenciamento em massa ou hibridação em microarranjos (GERBER; HERSCHLAG; BROWN, 2004; TENENBAUM et al., 2000; TOWNLEY-TILSON et al., 2006).

### 1.3 Tripanosomatídeos

Os tripanossomatídeos, termo adaptado do latim que designa a família Trypanosomatidae, compreendem os organismos da ordem Kinetoplastida, Filo Euglenozoa. São protozoários caracterizados por conter uma estrutura celular particular, o cinetoplasto, região da mitocôndria única que contém o kDNA, ou DNA do cinetoplasto (MASLOV; PODLIPAEV; LUKE, 2001).

A família Trypanosomatidae é caracterizada como heterotrófica de hábito parasitário obrigatório, contendo um único flagelo e um cinetoplasto pequeno. Parasitam todas as classes de vertebrados, alguns invertebrados e também plantas.

Algumas das espécies representantes são de grande importância sanitária, causadoras de graves doenças em animais e seres humanos, como a leishmaniose (*Leishmania* spp.), doença de Chagas (*Trypanosoma cruzi*) e doença do sono (*Trypanosoma brucei*).

O Filo Euglenozoa é um grupo divergente dos demais eucariotos, baseado tanto em análises filogenéticas comparando grupos especiais de genes, como também em características moleculares particulares. Primitivos, no sentido da ausência de vias comuns da regulação realizada por outros eucariotos; evoluídos, no sentido de serem tão bem adaptados ainda que com a falta de mecanismos moleculares básicos, driblando essa ausência através de mecanismos que ainda estão por serem descobertos (CLAYTON, 2002; PHILIPPE et al., 2000). Por esta razão, o conhecimento dos seus aspectos moleculares servem como referência ao estudo também em eucariotos superiores.

Os tripanossomatídeos abrigam várias destas particularidades que os tornam atraentes modelos de estudo na biologia celular e molecular dos organismos. O presente estudo dará enfoque nas particularidades do *T. cruzi*.

### 1.3.1 Biologia

O organismo *T. cruzi* é um parasita heteroxênico, pois ocupa dois hospedeiros de grupos distintos no seu ciclo de vida: um inseto hematófago e um vertebrado mamífero. O inseto é comumente conhecido por barbeiro, pertencente à ordem Hemiptera, família Reduviidae, subfamília Triatominae. Mais de 130 espécies estão na lista de transmissores dos parasitas, mas os mais comuns são: *Triatoma infestans*, *Triatoma brasiliensis*, *Triatoma dimidiata*, *Rhodnius prolixus*, e *Panstrongylus megistus* (MANOEL-CAETANO; SILVA, 2007). Entre os hospedeiros mamíferos estão roedores, animais domésticos, marsupiais, primatas e o homem (BRIONES et al., 1999).

Ocupando ambientes tão divergentes, sofre drásticas modificações morfológicas e bioquímicas no objetivo de sobreviver às condições ambientais. Sendo um parasita obrigatório, para sua continuidade necessita de uma capacidade replicativa e de se propagar para outros organismos. A replicação acontece quando se encontra em ambientes com abundância nutricional. Em geral, o estresse nutricional ou a densidade populacional desencadeiam processos de transição das formas de vida, quando passam a ser infectivas. Neste estágio, as células reduzem seu metabolismo ao mínimo e aumentam grandemente sua motilidade para garantir a infecção, ou então morrem em relativamente pouco tempo.

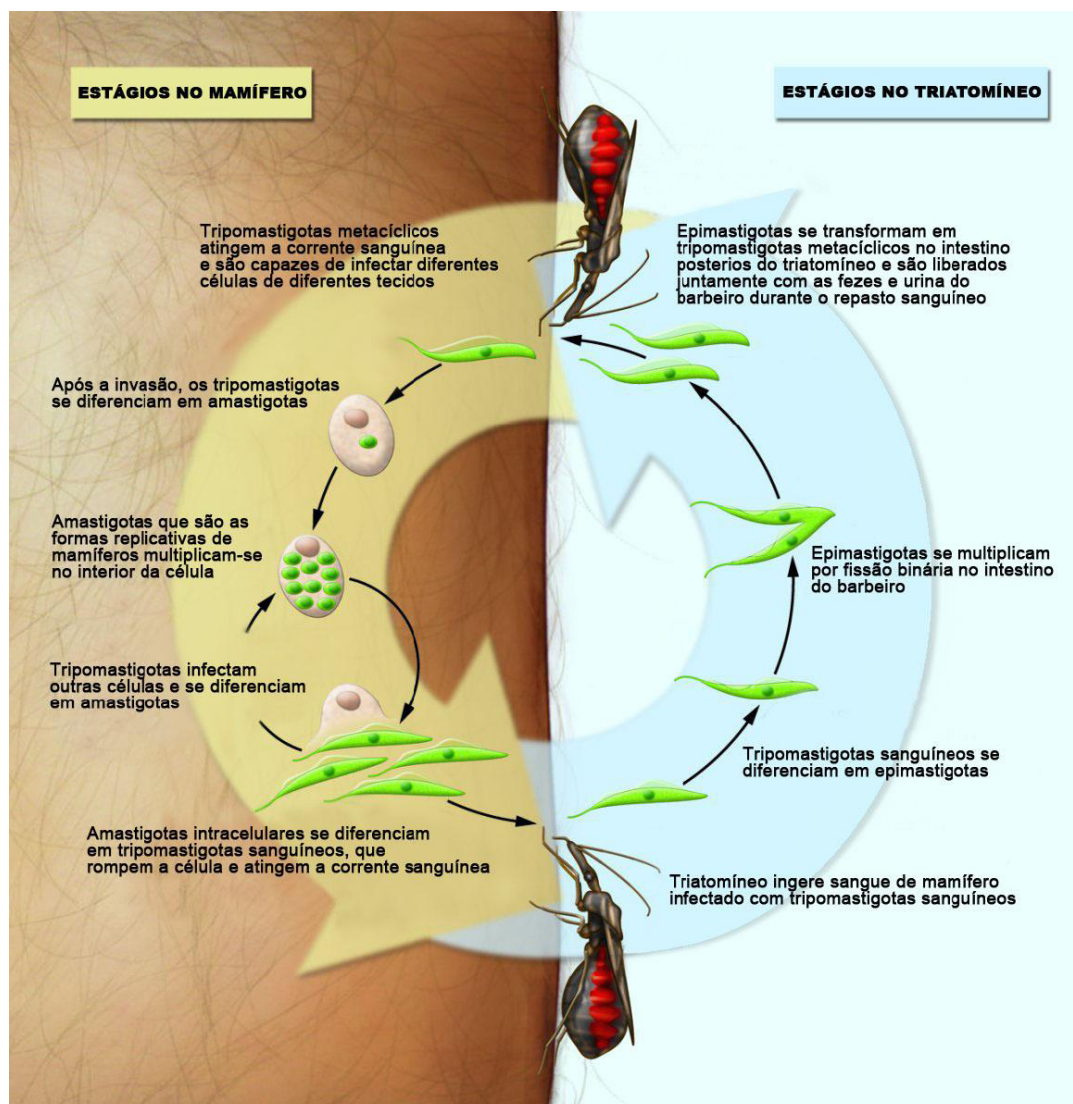
Há diversas cepas encontradas na natureza. Diferem na capacidade de invadir células nos hospedeiros, na parasitemia e na morfologia. As cepas são classificadas em grupos de acordo com o alinhamento de determinados genes. Várias classificações já foram propostas, sendo que a mais recente enquadra as cerca de 50 cepas em 6 grupos distintos, nomeados TcI a TcVI (ZINGALES et al., 2009).

### 1.3.2 Ciclo de vida

Para o organismo *T. cruzi*, são consideradas quatro formas distintas no seu ciclo de vida, apesar de serem vistas também várias formas transicionais. São elas: epimastigotas, formas replicativas do trato digestivo do inseto; tripomastigotas metacíclicas, formas infectivas que se desenvolvem no final do intestino do inseto; amastigotas, formas replicativas e também infectivas das células do mamífero;

tripomastigotas sanguíneas, da corrente sanguínea do mamífero (CHAGAS, 1909, 1935; DE SOUZA, 2002).

As formas tripomastigotas sanguíneas do hospedeiro mamífero entram pelo aparelho bucal do inseto durante seu repasto sanguíneo. Os parasitas migram até a porção inicial do intestino. Num ambiente pobre em monossacarídeos, com temperatura inferior e pH mais ácido, a maioria dos parasitas diferencia nas formas epimastigotas (TYLER; ENGMAN, 2001). Ali se proliferam quando há nutrientes oriundos de uma nova alimentação. Ao final do intestino, onde há escassez nutricional, aderem pelo flagelo à ampola retal para iniciar a metaciclogênese, processo de diferenciação para formas tripomastigotas metacíclicas infectivas. Dali se desprendem da superfície de adesão e saem juntamente com as excretas do inseto (GARCIA; GONZALEZ; AZAMBUJA, 1999). Os tripomastigotas adentram no mamífero por discontinuidades na epiderme ou mucosas, podendo invadir células nucleadas. O mecanismo de entrada nas células mais conhecido é através de uma invaginação de membrana do hospedeiro, promovida pela reorganização do citoesqueleto na região da invasão na membrana plasmática e recrutamento de proteínas específicas da célula do mamífero (ANDRADE; ANDREWS, 2004; REIGNAULT et al., 2014). É formado um vacúolo parasitóforo que envolve todo o parasita até que este se encontre totalmente dentro da célula. Ali ele inativa a via do sistema imune que sinalizaria a infecção e provocaria a morte celular. Neste ambiente ácido, provocado pelas enzimas oriundas de lisossomos recrutados, o parasita permanece de 2 a 4 horas, quando deve romper a membrana do vacúolo, ou senão morre. No citoplasma sofre um novo processo de diferenciação, a amastigogênese, cujas formas amastigotas adotam morfologia esférica e reentram no ciclo celular para proliferar. Quando a célula hospedeira já estiver repleta de parasitas, inicia-se nova diferenciação para uma forma epimastigota alongada intermediária ao tripomastigota. Estes podem romper a célula por ação de proteases ou mesmo pelos próprios batimentos do flagelo. Ao escapar da célula, os tripomastigotas sanguíneos são capazes de invadir novas células e, ao entrar na corrente sanguínea, espalhar-se para outros tecidos. Na corrente, podem ser ingeridos pelo triatomíneo, reiniciando o ciclo (Fig. 4) (TYLER; ENGMAN, 2001).



**Figura 4.** Ciclo de vida do *T. cruzi*. Fonte: NARDELI, 2009, modificado de (STUART et al., 2008).

### 1.3.3 Biologia celular

O *T. cruzi* apresenta características celulares próprias que o diferenciam dos outros organismos unicelulares.

É observada uma rígida estrutura de microtúbulos, o citoesqueleto subpelicular, localizado logo abaixo da membrana plasmática, o qual confere alta resistência contra ruptura mecânica (DE SOUZA, 2002). Seu flagelo emerge de uma invaginação endocítica denominada bolsa flagelar. Esta estrutura é responsável pela endo e exocitose que ocorre no parasita, uma vez que a camada de microtúbulos impede a



formação de invaginações na membrana. Os microtúbulos subpeliculares estão ausentes apenas na região da bolsa flagelar.

Os glicossomos estão distribuídos por todo o citoplasma; são semelhantes aos peroxissomos em outros eucariotos, porém contém todas as enzimas glicolíticas, as quais em outros organismos se encontram dispersas no citoplasma (DE SOUZA, 2002).

Possui apenas uma mitocôndria, que se estende ao longo de toda a célula. Esta mitocôndria apresenta uma região especializada, denominada cinetoplasto. O cinetoplasto é uma estrutura particular da ordem. Contém um concatenado de DNAs circulares, arranjados na matriz da mitocôndria única, e proteínas associadas envolvidas com a manutenção da estrutura e a replicação do cinetoplasto. Sua localização é alterada durante o ciclo de vida, permitindo caracterizar o estágio em que o parasita se encontra (MORRIS et al., 2001). Nas formas replicativas as moléculas de kDNA são bem empacotadas, formando um disco côncavo, e orientadas paralelamente ao eixo longitudinal do parasita. Podem aparentar ter uma dupla camada ao microscópio óptico. Em tripomastigotas o cinetoplasto é subterminal (DE SOUZA, 2002).

O material genético do *T. cruzi* está concentrado em um núcleo que sofre rearranjos durante o ciclo de vida. Enquanto na fase proliferativa, o núcleo é arredondado, com nucléolo evidente e pouca heterocromatina periférica. Quando diferencia para as formas infectivas, a forma do núcleo se torna alongada, quase todo tomado por heterocromatina e o nucléolo se desfaz (ELIAS et al., 2001).

#### 1.3.3.1 Reservossomos

Os reservossomos são grandes organelas especializadas, exclusivas de *T. cruzi*, presentes na região posterior das formas epimastigotas. São variáveis em tamanho e número, mas em geral possuem entre 5-10 unidades com cerca de 400-600 nm (SOARES; DE SOUZA, 1988; SOARES; SOUTO-PADRÓN; DE SOUZA, 1992).

São conhecidas por armazenamento nutricional, já que é o destino final de vesículas endocitadas através do citóstomo (DE SOUZA, 2009). Inicialmente foram descritas como sendo organelas lisossomos-like, por estarem também realizando

degradação de macromoléculas e síntese de lipídeos (SANT'ANNA et al., 2008a). Além disso, foi reportada atividade autofágica nestas estruturas (SANT'ANNA et al., 2008b).

Com novas tecnologias de identificação de ácidos nucleicos e proteínas, é possível que estas organelas reservem ainda outras funções desconhecidas, as quais podem abrigar importantes segredos das particularidades do ciclo de vida parasita.

### *1.3.4 Biologia molecular de T. cruzi*

#### 1.3.4.1 Genoma

O genoma nuclear do *T. cruzi* contém cerca de 32,5 Mb, organizados em 41 pequenos cromossomos. A estrutura dos cromossomos foi determinada anos após seu sequenciamento completo (WEATHERLY; BOEHLKE; TARLETON, 2009), devido a dificuldades na montagem no primeiro sequenciamento (EL-SAYED et al., 2005), iniciado em 1997 (ZINGALES et al., 1997).

É constituído por cerca de 10 mil genes (segundo TryTripDB 8.0). É um genoma bem compacto, quase totalmente ausente de *introns*. De fato, a única demonstração de *cis*-splicing foi para o gene codificante da Poli(A) polimerase (MAIR et al., 2000). Em *T. brucei* foi encontrado um *intron* em um isotipo de gene de tRNA (TAN et al., 2002). Os genes são bem próximos, com regiões intergênicas curtas. Estão arranjados em unidades policistrônicas, sem relação funcional entre seus produtos (CAMPBELL; THOMAS; STURM, 2003). As unidades transcricionais contêm de dezenas a centenas de genes e podem estar localizadas em ambas as fitas. Entre as unidades podem haver arranjos sequenciais de tRNA (MARTÍNEZ-CALVILLO et al., 2010).

#### 1.3.4.2 Transcrição

Não há regulação individual da transcrição dos genes de *T. cruzi*. Assim, todos os genes do *T. cruzi* são transcritos concomitantemente, sendo regulados apenas a partir do processamento em transcritos individuais, quando são então direcionados para

tradução ou degradação, determinando assim o perfil de expressão diferencial dos genes.

O primeiro promotor encontrado para a RNA polimerase II sinalizando o início da transcrição está na sequência do gene do mini-éxon (GILINGER; BELLOFATTO, 2001). Em teoria, há um promotor no início das unidades, possivelmente uma região rica em As e Ts (CUROTTO DE LAFAILLE; LABAN; WIRTH, 1992), o que ainda não foi provado. Outros autores argumentam que há apenas uma região de início da transcrição no cromossomo e que a transcrição segue até o final do cromossomo (MARTÍNEZ-CALVILLO et al., 2003). Há argumentos suportando a ideia de que a transcrição seria randômica e a polimerase se associaria em regiões de baixa afinidade, sendo favorecida assim por acessibilidade genômica e porções de anelamento/desnaturação da dupla-fita (MCANDREW et al., 1998).

Ao longo do ciclo de vida do parasita, a transcrição sofre alterações globais. A atividade da RNA Polimerase II é reduzida nos tripomastigotas metacíclicos, estando ainda ativa nas formas intermediárias da metaciclogênese (FERREIRA et al., 2008).

#### 1.3.4.3 Processamento pós-transcricional

O pré-mRNA policistrônico contém as sequências de diversos genes, sem relação de função ou mesmo de regulação entre seus produtos (VANHAMME; PAYS, 1995). Esses transcritos devem, portanto, ser segregados a fim de que cada um possa seguir seu destino na célula e sofrer regulação individual.

No processo de *trans-splicing*, as unidades codificadoras são separadas e é incorporado à região 5' dos transcritos primários o mini-éxon, ou *spliced-leader* (SL), proveniente de outro mensageiro, o qual contém a metil-guanosina trifosfato (m7Gppp). Este segmento contém 39 a 41 bases e tem como principal função sinalizar o mRNA para a tradução. Esta sequência é inserida por uma maquinaria que reconhece um sinal AG precedido por uma região rica em pirimidinas no transcrito primário policistrônico. Esta maquinaria de *trans-splicing* está acoplada a componentes que realizam adição da cauda poli(A). Elas realizam simultaneamente a adição da sequência SL na região 5' do transcrito à jusante desta região sinalizadora, e

a clivagem e adição da cauda poli(A) na região 3' do transcrito à montante (HAILE; PAPADOPOULOU, 2007; LEBOWITZ et al., 1993; LIANG et al., 2003).

#### 1.3.5 Regulação da expressão gênica em *T. cruzi*

O organismo *T. cruzi*, juntamente com outros kinetoplastídeos, apresenta várias atribuições e especializações incomuns ou únicas. Não há regulação transcricional individual clara nestes organismos, visto a ausência de promotores canônicos para RNA polimerase II definidos (HAILE; PAPADOPOULOU, 2007). A regulação transcricional é global, já que os fatores transcricionais são gerais. Em tripomastigotas sanguíneos, a atividade da RNA polimerase II é diminuída, o que reflete na redução geral da transcrição (ELIAS et al., 2001). Os genes adjacentes são transcritos em unidades policistrônicas, as quais não apresentam funções relacionadas ou mesmo padrão regulatório em comum (ABUIN et al., 1999). Estas unidades são processadas por *trans-splicing*, sendo as regiões 5' e 3' do transcrito processadas simultaneamente. (CLAYTON, 2002).

A expressão estágio-específica de genes que estão no mesmo policistron e a ausência de relação funcional evidente entre os produtos dos transcritos indicam que o controle da expressão gênica é essencialmente pós-transcricional. Assim, há possibilidade de regulação nas etapas de *trans-splicing*, exportação do núcleo para o citoplasma, degradação citoplasmática dos mRNAs, associação à maquinaria de tradução e estabilidade protéica. Estudos investigam a estabilidade do mRNA e níveis de tradução como evidência desta proposição (AVILA et al., 2003; CLAYTON; SHAPIRA, 2007).

Durante os processos de diferenciação em *T. brucei*, 25% dos genes que sofrem aumento na expressão codificam para proteínas envolvidas na interação com o RNA e na tradução. Estes transcritos de RBPs, por serem a resposta imediata a estímulos de diferenciação, podem ser considerados equivalentes aos “fatores transcricionais primários” dos eucariotos superiores, porém pertencendo ao nível pós-transcricional. Mais uma vez é ressaltado o fato de serem organismos que dependem majoritariamente da regulação pós-transcricional da expressão gênica para rapidamente responder aos estímulos recebidos (KRAMER, 2012).

As regiões não traduzidas (UTRs) nos mRNAs abrigam a informação da sua destinação regulada. Proteínas especializadas se associam a estas regiões, que podem adotar conformações estruturais, tornando-as mais particulares e diferenciadas. O tamanho destas regiões varia entre os organismos, mas é observado um padrão dentro de cada organismo. Em *T. cruzi*, uma análise global apontou que a região 5' UTR dos mRNAs têm em média apenas 35 nt e a 3' UTR 264 nt (CAMPOS et al., 2008).

Há diversos elementos caracterizados nas UTRs de eucariotos. Cada elemento corresponde a um sinal específico reconhecido por proteínas que irão desempenhar sua função designada. É importante que os elementos que sinalizam destinos diferentes para o transcrito sejam distintos, a fim de que o processo de regulação seja fino (MIGNONE et al., 2002). Duas RBPs da família RRM em *T. cruzi*, TcUBP1 e TcRBP3 tiveram seu motivo de interação no mRNA identificado (um deles sendo estrutural), e estão enriquecidos em grupos específicos de genes (NOÉ; GAUDENZI; FRASCH, 2008). Em *T. brucei* são observados diversos motivos conservados em transcritos cujos produtos participam de um mesmo processo celular (MAO; NAJAFABADI; SALAVATI, 2009), o que remete à já comprovada teoria dos operons pós-transcricionais. Uma vez que sejam identificadas as proteínas que se ligam a estes motivos nos tripanosomatídeos e suas modificações pós-traducionais sejam mapeadas, será possível determinar como as decisões do destino dos RNAs estão ligadas às condições do ambiente em que o parasita se encontra (CLAYTON; SHAPIRA, 2007).

Em suma, este estudo é justificável em *T. cruzi*, sendo um organismo de regulação da expressão gênica estritamente em nível pós-transcricional, dependendo assim de outros mecanismos para controlar a expressão de genes estágio-específicos frente aos tão distintos ambientes que ocupa. E não apenas justificável, mas se trata do modelo ideal no estudo da coordenação das redes de expressão, já que interferências celulares a fim de estudar o sistema terão uma resposta estritamente no nível pós-transcricional, e não proveniente de alterações na taxa de transcrição de genes específicos (CAMPBELL; THOMAS; STURM, 2003).

### 1.3.6 Domínio RRM em *T. cruzi*

Há ainda poucos estudos notáveis acerca do envolvimento de RRM's no controle da abundância e da tradução do mRNA em tripanosomatídeos. A maioria destas RBPs ainda não possui função biológica definida.

Dentre as primeiras proteínas da família RRM a serem estudadas estão as proteínas TcUBP1 e TcUBP2. A TcUBP1 reconhece um elemento no mRNA de instabilidade rico em AU nos transcritos de mucinas, direcionando-as para degradação. Tem semelhanças com a proteína ELAV humana, porém com efeito antagonista sobre o RNA. Sua região C-terminal, rica em glicina, permite homodimerização, ou heterodimerização com a proteína TcUBP2, formando um complexo de regulação negativa. Esta ligação é rompida pela proteína de união a poli(A) (PABP) (D'ORSO; FRASCH, 2001a, 2001b, 2002). Os transcritos alvos da TcUBP1 foram identificados anos mais tarde. A maior parte deles codifica para proteínas envolvidas no metabolismo das glicoproteínas. Também foram caracterizados os alvos da TcRBP3, cuja maior parte codifica para subunidades ribossomais. O elemento de reconhecimento destas duas proteínas foi identificado e mostrou envolver a estrutura secundária dos RNAs aos quais se ligam (NOÉ; GAUDENZI; FRASCH, 2008). Foi postulado que a TcUBP1 possa estar envolvida no metabolismo de mRNAs em ambos os lados do complexo do poro nuclear, servindo o RRM como sinal de localização nuclear e sendo transportada associada ao RNA (CASSOLA; FRASCH, 2009).

A proteína TcRBP19 é uma proteína de baixa expressão presente em amastigotas, e reconhece preferencialmente homorribopolímeros de poli(C). Recentemente foi visto que esta RBP sofre um processo de auto-regulação por *feedback* negativo, se ligando à 3' UTR do próprio mRNA. Além disso, a superexpressão da proteína afetou a metaciclogênese e a capacidade de infecção dos parasitas, ressaltando a importância da família para o desencadeamento de processos importantes na biologia do organismo (PÉREZ-DÍAZ et al., 2007, 2012, 2013).

### 1.3.7 TcRBP40

A proteína TcRBP40 foi identificada em um estudo anterior à presente tese, a partir de uma busca por proteínas contendo o domínio RRM em *T. cruzi* (GUERRA-SLOMPO et al., 2012). Esta nova anotação de proteína de união ao RNA para a TcRBP40, anteriormente tida por proteína hipotética, se acrescentou à importância de ser uma RBP exclusiva de *T. cruzi*, não compartilhada pelos outros Kinetoplastídeos. Isso não é comum, visto que de 81 proteínas contendo RRM, apenas outras 2 seriam exclusivas do organismo (Fig. 5).



**Figura 5.** Número de RBPs contendo RRM nos TriTryps. Gráfico de Venn representativo do número de RBPs contendo o domínio RRM putativo em *T. cruzi* (amarelo), *T. brucei* (verde) e *L. Major* (roxo). Valores de *T. cruzi* a partir de GUERRA-SLOMPO et al., 2012, e dos demais de acordo com GAUDENZI; FRASCH; CLAYTON, 2005.

A investigação da proteína TcRBP40 mostrou que ela tem expressão diferencial ao longo do ciclo de vida do parasita, sendo expressa em epimastigotas e ausente na forma metacíclica, o que foi suportado por *Western blot* e imunofluorescência. Uma análise ribonômica de captura *in vitro* dos mRNAs associados e identificação por microarranjo mostrou que a proteína TcRBP40 recombinante identifica alvos de maneira seletiva, mostrando um enriquecimento de alvos contendo elementos ricos em AG na sua região 3' UTR. A localização celular por imunofluorescência mostrou que a TcRBP40 co-localiza com a proteína cruzipaína, sugerindo a sua presença nos reservossomos do parasita. Experimentos adicionais foram realizados para confirmar esta localização celular, muito incomum considerando-se uma organela de estocagem nutricional. Desta análise trata a primeira seção dos resultados desta tese.

Adicionalmente, foi observada uma alta identidade entre a proteína TcRBP40 e outra proteína contendo o domínio RRM em *T. cruzi*, a TcRBP7. A segunda parte desta tese visa a análise filogenética destas duas proteínas em tripanossomatídeos, verificando a existência de uma relação de homologia entre os dois genes codificantes. Uma vez determinada esta relação, seguiu-se à comparação funcional entre as duas proteínas, visto tratarem-se de elementos regulatórios em nível pós-transcricional em um organismo modelo para este estudo. Este modelo poderá servir para melhor compreender o comportamento das redes regulatórias da expressão gênica diante da duplicação de fatores que as coordenam.



## 2 OBJETIVO

### 2.1 Geral

Caracterizar as proteínas regulatórias da família RBP7 em *T. cruzi* e *T. brucei*, buscando relacionar as divergências genéticas com os seus aspectos funcionais.

### 2.2 Objetivos Específicos

- Clonar e expressar as proteínas TcRBP7 e TcRBP40 na Plataforma *Gateway*, Invitrogen e recombinar os genes de interesse a vetores para expressão em sistemas bacterianos, e vetores para transfecção em *T. cruzi* para ensaios de purificação por afinidade e superexpressão;
- Caracterizar as proteínas quanto ao seu perfil de expressão no ciclo de vida e localização celular;
- Analisar filogeneticamente a proteína TcRBP40 de *T. cruzi* com as proteínas homólogas de outros tripanossomatídeos, inferindo sua origem;
- Determinar a estrutura tridimensional das proteínas por modelagem molecular por homologia.
- Determinar os mensageiros alvos das proteínas de *T. cruzi* por sequenciamento em larga escala, obtidos da purificação de complexos RNP *in vivo*.

### 3 MATERIAIS E MÉTODOS

#### 3.1 Tampões, soluções e meios

PBS: NaCl 137 mM, KCl 2,7 mM, Na<sub>2</sub>HPO<sub>4</sub> 4,3 mM, KH<sub>2</sub>PO<sub>4</sub> 1,5 mM.

Meio LIT: Infuso de fígado (Difco) 0,5%, Bacto-triptose (Difco) 0,5%, Soro fetal bovino (Cult Lab) 10%, Hemina 0,0025%, Glicose 10 mM, NaCl 75,3 mM, KCl 5,4 mM, Na<sub>2</sub>HPO<sub>4</sub> 56,4 mM.

TM: Tris-HCl 20 mM pH 7,2, MgCl<sub>2</sub> 2 mM.

TA: NaH<sub>2</sub>PO<sub>4</sub> 28 mM, Na<sub>2</sub>HPO<sub>4</sub> 71 mM, SDS 1%, β-mercaptoetanol 1%, Uréia 6 M, Azul de bromofenol 0,015%.

Tampão para transferência de proteínas: Tris-base 25 mM, Glicina 192 mM, Metanol 20%.

Tampão de eletroporação: NaCl 140 mM, Hepes ácido 25 mM, Na<sub>2</sub>HPO<sub>4</sub> 0,74 mM.

Tampão Zeiner (FISH): Solução de Denhardt 10x (Sigma), SSC 4x, EDTA 1 mM, formamida deionizada 35%.

SSC 20x: NaCl 3 M, C<sub>6</sub>H<sub>5</sub>Na<sub>3</sub>O<sub>7</sub> 0,3 M.

IMP0: HEPES 10 mM pH 7,4, KCl 100 mM, MgCl<sub>2</sub> 5 mM.

IMP1: HEPES 10 mM pH 7,4, KCl 100 mM, MgCl<sub>2</sub> 5 mM, NP-40 0,5%.

IMP2: HEPES 10 mM pH 7,4, KCl 100 mM, MgCl<sub>2</sub> 5 mM, NP-40 1%.

TBE: Tris-base 89 mM, Ácido bórico 89 mM, EDTA 2 mM.

TE: Tris-HCl pH 8,0 10 mM, EDTA 1 mM.

Meio LB: Bacto-triptona 1%, extrato de levedura 0,5%, NaCl 85,6 mM.

Tampão de lise solúvel para *E. coli*: Tampão fosfato de sódio 10 mM pH 6,8 ou Tris-HCl 10 mM pH 9,5, NaCl 300 mM, glicerol 10%, NP-40 0,5%.

Tampão de ligação para RNA *pull-down*: Tris-HCl 10 mM pH 8,0, MgCl<sub>2</sub> 10 mM, KCl 10 mM, DTT 0,1 mM.

## 3.2 Preparações com culturas de *T. cruzi*

### 3.2.1 Cultivo dos parasitas

Formas epimastigotas de *Trypanosoma cruzi*, cepa Dm28c, clone CR9.1 (Contreras, 1988), foram cultivadas em meio LIT contendo 20% de soro fetal bovino e penicilina e streptomomicina a uma concentração de  $_{\text{ug/mL}}$ , em frascos de vidro, garrafas de cultivo ou tubos Falcon, em estufa a 28 °C.

Todos os ensaios foram realizados com culturas em fase exponencial final de crescimento, quando a cultura atingia uma densidade aproximada de  $3 \times 10^7$  células/mL.

### 3.2.2 Extratos proteicos de células

O lisado total de proteínas foi obtido a partir de  $5 \times 10^8$  parasitas, centrifugados e lavados com PBS (Tampão fosfato-salino), ressuspensando em 500  $\mu\text{L}$  de PBS e TA (Tampão de Amostra) para uma concentração de  $10^6$  células/L.

O lisado solúvel, ou fração livre de organelas, foi obtido a partir de  $10^9$  parasitas, lavados e incubados com 1 mL de tampão de lise IMP1 por 2 h, a 4 °C, sob agitação leve. Após esse período, o extrato foi centrifugado a 30.000 xg por 30 min, a 4 °C. Uma fração foi mesclada com TA 4x, submetida a agitação vigorosa por aprox. 2 min, incubada a 95 °C por 5 min e centrifugada por 1 min a 10.000 xg.

A fração pesada, contendo núcleo e organelas, foi obtida a partir do precipitado do lisado acima mencionado. Foi ressuspensado em 1 mL de PBS e tratado com TA 4x nas mesmas condições.

A fração enriquecida de reservossomos para o *western blot* foi preparada a partir do precipitado final da purificação, ressuspensando diretamente em TA 4x na última gota do tubo. O volume total resultante foi aplicado no gel, a fim de obter proteínas suficientes para visualização no ensaio.

### 3.2.3 Lise solúvel de *T. cruzi*

Um total de  $10^9$  parasitas das culturas transfectantes de *T. cruzi* com TcRBP7-GFP, TcRBP40-GFP ou GFP-controle foram submetidas a lise solúvel com tampão hipotônico e detergente não-iônico.

As culturas de 3 dias, em fase exponencial de crescimento, foram submetidas a 10 min de centrifugação a 5.000 xg, a 4 °C, lavadas duas vezes com tampão PBS e centrifugadas nas mesmas condições. Por fim, foram suspensas em tampão de imunoprecipitação IMP1 (HOLETZ et al., 2010), acrescido de coquetel inibidor de proteases (cOmplete EDTA-free, Roche) e inibidor de RNAses (40 U de RNase OUT, Invitrogen), e lisadas por 2 horas sob agitação orbital. Todo o tratamento das amostras foi realizado a 4 °C. O sobrenadante solúvel foi obtido de uma centrifugação a 20.000 xg por 20 min e separado do precipitado.

As frações foram analisadas em *Western blot* para verificar a expressão e a presença das fusões na fração desejada, a saber, a fração solúvel.

### 3.2.4 Transfecção em *T. cruzi*

$4 \times 10^7$  parasitas na forma epimastigota foram lavados em PBS e ressuspensos em tampão de eletroporação, onde incubaram por 10 min em gelo. Foram transferidos para cubeta de eletroporação de 0,2 mm, juntamente com 15-20 µg de vetor em volume máximo de 50 uL, e submetidos a dois pulsos de aprox. 4 ms a 450 V, 500 µF. Incubaram mais 10 min em gelo e foram transferidos para garrafa de cultura contendo 4 mL de meio LIT completo. Após 24 h foi adicionado 250 µg/mL de neomicina (G418, Sigma), e após 48 h novamente adicionado para uma concentração final de seleção de 500 µg/mL. Nos primeiros 7 dias foram realizados repiques na proporção de 1:4 quando a cultura atingia alta quantidade de parasitas, mantendo a concentração da droga. Após a interrupção do crescimento, as culturas permaneceram de 15-21 dias para seleção, confirmada pela morte da cultura controle, sem o vetor.

A expressão da proteína fusionada foi verificada por microscopia de fluorescência ou *western blot* de extrato protéico dos transfectantes.

### 3.2.5 Purificação de reservossomos

Para obter uma fração celular enriquecida de reservossomos, um protocolo já estabelecido foi utilizado (CUNHA-E-SILVA et al., 2002; SANT'ANNA et al., 2008b). Assim,  $10^{10}$  parasitas foram centrifugados a 5.000 xg por 10 minutos a 4 °C, lavados duas vezes com tampão TM com sacarose 250 mM na mesma condição e ressuspensos em 13 mL do mesmo tampão. Os parasitas foram submetidos a 20 pulsos de 2 segundos de sonicação a 20% em gelo, com intervalos de 1 s. Foi possível observar ao MO que a maioria dos parasitas foi lisado, restando apenas fragmentos celulares. O lisado foi centrifugado a 2.450 xg por 10 min a 4 °C e o sobrenadante mesclado em igual volume de TM com sacarose 2,3 M, resultando em uma concentração de 1,27 M de sacarose. Foram aplicados 12 mL deste lisado no fundo de um tubo de 50 mL de ultracentrifugação. Acima dele, foi adicionado 10 mL de TM com sacarose 1,2 M, em seguida 8 mL com sacarose 1 M, e então 5 mL com sacarose 0,8 M. Este gradiente foi submetido a ultracentrifugação a 97.000 xg por 2 h 30 min, a 4 °C, em um rotor P28S (ultracentrífuga Hitachi). A amostra foi observada em 3 regiões distintas: a primeira na fração entre 0,8 M e 1 M, a segunda entre 1 M e 1,2 M e uma terceira ao fundo do tubo. As frações superiores foram coletadas e a elas foi adicionado TM para obter uma concentração de sacarose de 250 mM. Foram submetidas a ultracentrifugação a 120.000 xg por 30 min a 4 °C, no rotor P40ST. O precipitado foi ressuspensionado em tampão apropriado para ensaios de proteínas ou lise de reservossomos. A amostra da fração entre 1,2-1,0 M foi caracterizada anteriormente por microscopia eletrônica como sendo enriquecida de reservossomos, e foi a utilizada para os ensaios subsequentes.

### 3.2.6 Ensaio de endocitose

$10^7$  epimastigotas foram lavados e incubados em PBS por 15 min a 28 °C, a fim de intensificar a endocitose após este período. Foi então adicionada transferrina conjugada a Alexa 533 a uma concentração de 100 µg/mL, permitindo as células incorporarem a molécula por 30 min a 28 °C. Após, foram lavados 3 vezes em PBS e preparados para microscopia de fluorescência. Esta metodologia foi anteriormente descrita por (EGER; SOARES, 2012).

### 3.3 Imunoensaios

#### 3.3.1 *Western blotting*

Neste ensaio, frações de proteínas separadas por eletroforese em SDS-PAGE 15% (SAMBROOK; RUSSEL, 2001) foram transferidas para uma membrana de nitrocelulose Hybond-C (Amersham), por eletroforese horizontal, em tampão específico, por 2 h a 60 V ou 1 h a 100 V. A membrana foi inicialmente corada com Ponceau S para visualização da transferência e marcação de posição. O bloqueio foi feito por pelo menos 1 h em 5% de leite desnatado diluído em PBS acrescido de Tween 20 (PBS-T) 0,1%. A membrana foi lavada 3 vezes em PBS-T, sob agitação por 5 min, e incubada por 1h com os anticorpos específicos, diluídos em PBS-T nas seguintes titulações: soro policlonal anti-TcRBP40 a 1:200, soro policlonal anti-GAPDH a 1:500, e soro policlonal anti-H2A.z a 1:1000. Os soros policlonais foram produzidos em projetos anteriores (projeto de Mestrado anterior a este, ou gentilmente cedidos por Stenio P. Fragoso e Gisele Picchi, respectivamente – dados não publicados). O anticorpo anti-cruzipaína, gentilmente cedido pelo Dr. Maurílio J. Soares (BATISTA et al., 2014), foi produzido a partir de cultura clonal de hibridoma, onde o próprio sobrenadante da cultura foi utilizado na incubação com a membrana. Após a incubação do anticorpo primário, a membrana foi lavada 3 vezes e incubada 1 h com anticorpo secundário anti-IgG de camundongo conjugado a fluoróforo DyLight 680 nm ou 800 nm, ou conjugado à fosfatase alcalina. Após lavar 3 vezes, as associações proteína-anticorpo foram detectadas no sistema Odyssey (Li-COR), ou através da reação colorimétrica na presença de 5-bromo-4-cloro-3-indolil-fosfato e nitroazul de tetrazólio, de acordo com o conjugado no anticorpo secundário.

#### 3.3.2 *Microscopia de fluorescência*

Parasitas selvagens ou transfectantes foram lavados e fixados em solução com paraformaldeído 4%, diluído em PBS, em uma proporção de  $2 \times 10^6$  células para cada ensaio. Após 3 lavagens com PBS e centrifugação por 5 min a 5.000 xg, foram aderidos por 10 min em lâminas de fluorescência, de 3 ou 8 poços, tratadas com poli-L-lisina.

Nos ensaios de imunofluorescência indireta, as células foram permeabilizadas com Triton X-100 0,5% por 5 min, lavadas 5-6 vezes com PBS e bloqueadas com BSA 4% diluído em PBS, por 16 h, a 4 °C. Os anticorpos foram incubados em PBS-BSA por 1 h a 37 °C, e os poços em seguida lavados 3 vezes com PBS. Os anticorpos secundários conjugados com Alexa 488 ou 594 diluídos 1:400 foram incubados e lavados nas mesmas condições.

Os ensaios de hibridização *in situ* (FISH) foram realizados conforme anteriormente descrito (ALVES et al., 2013), com algumas modificações. As células foram fixadas e permeabilizadas conforme no ensaio de imunofluorescência, mas bloqueadas com tampão Zeiner completo, por 1 h à temperatura ambiente. As sondas de DNA fita simples marcadas com Cianina-5 na extremidade 5' foram utilizadas a uma concentração de 1 ng/μL, diluídas no mesmo tampão Zeiner, inicialmente desnaturadas a 70 °C por 3 minutos, e então incubadas a 37 °C por pelo menos 16h. Após, foram lavadas 5 vezes com SSC 2x. A sequência das sondas foi (5'-3'): sentido CTGCAAGTGTCTGTAAGATCGGAATTGATTTCTCCCTTTGGATGG CAACAGGCG, e anti-sentido GACGTTACAGCATGACACTTTCTAGCCTT AACTAAAGAGGGAAACCTACCGTTGTCCGC.

Para visualização do núcleo e cinetoplasto, os parasitas foram tratados com DAPI a 1 μg/mL por 15 min e lavados 2 vezes com PBS. Aos poços foi adicionado o anti-*fading* N-propil-galato e cobertos com lamínula.

As lâminas foram visualizadas em microscópio óptico Nikon E80i. Quando indicado, foi utilizado um software para tratamento de deconvolução das imagens, permitindo agrupar a fluorescência dispersa para o seu ponto de origem.

### 3.3.3 Imunoprecipitação

As frações solúveis das culturas transfectantes TcRBP7-GFP, TcRBP40-GFP e GFP-controle foram submetidas a ensaio de captura por afinidade da etiqueta fusionada, em triplicata biológica (transfecções independentes).

Para cada cultura, 50 μL de solução de esferas magnéticas conjugadas a IgG de coelho (Anti-rabbit IgG Dynabeads, Life Technologies) foram lavados com tampão IMP0 e incubados por 1 hora com anti-GFP (anti-GFP monoclonal rabbit antibody,

Invitrogen) e inibidor de RNase, sob agitação leve, em volume final de 500 µL de IMP0. As esferas foram lavadas com o tampão IMP1 e incubadas com o lisado solúvel das culturas transfectantes por 2 horas sob agitação orbital a 4 °C. O volume contendo material não ligado foi coletado e utilizado nas análises seguintes. As esferas foram lavadas 3 vezes com 1 mL de tampão IMP2 por 10 minutos, sob agitação leve.

O RNA contido nas esferas e na fração não-ligada foram extraídos utilizando o RNeasy Mini RNA Extraction Kit (Qiagen), de acordo com especificações do fabricante, e quantificados no sistema NanoDrop 1000 (Thermo Scientific).

### **3.4 Construção de vetores**

#### *3.4.1 Amplificação dos genes*

Os genes codificantes das proteínas TcRBP7, TcRBP40, TbRBP7A e TbRBP7B foram selecionados a partir da sequência depositada no banco de dados TriTrypDB. Com o programa PrimerSelect, do pacote DNASTAR, oligonucleotídeos iniciadores foram desenhados para amplificar os genes das proteínas em estudo. O Quadro 1 contém o número de identificação dos genes, as sequências dos iniciadores, a temperatura de anelamento ideal e o tamanho dos produtos da amplificação. Na extremidade 5' de cada iniciador foi adicionado a sequência de recombinação *attB* no iniciador sentido (GGGGACAAGTTTGTACAAAAAAGCAGGCTTC) e anti-sentido (GGGGACCACTTTGTACAAGAAAGCTGGGTC).



**Quadro 1.** Iniciadores para os genes em estudo.

Proteína	ID do gene	Sequência 5'-3'	Tm	Gene (pb)
TcRBP7	TcCLB.506565.4	ATGGTGCTGGGACAGGCACG	62	381
		TCACGTGTGGCACTGTTTCGTTTCAGG	61,8	
TcRBP40	TcCLB.506565.12	ATGCCACAATCGAAACCAAGAGCC	60,3	357
		TCACAACAAACGTTTGGCTGGAGTG	60,9	
TbRBP7A	Tb927.10.12090	ATGCCACCGCGGGCTCG	62,4	351
		TCACCGTTGAACACGTTGTGGTG	60,2	
TbRBP7B	Tb927.10.12100	ATGCCACCGCGGGCTCG	62,4	351
		TCAACGTTGCACACGTTGTGGTG	61,2	

ID: identificação de acesso na base de dados

Tm: Temperatura de anelamento em °C

pb: pares de bases

F: iniciador sentido; R: iniciador anti-sentido

A reação de amplificação foi realizada por reação em cadeia da polimerase (PCR) utilizando a enzima *Taq* DNA Polimerase (Produzida pelo Insituto de Biologia Molecular do Paraná – IBMP). 100 ng de DNA genômico, extraídos pelo método de precipitação com fenol-clorofórmio (SAMBROOK; RUSSEL, 2001), foram submetidos a 5 min de desnaturação inicial a 94 °C, seguidos de 30 ciclos de desnaturação a 94 °C por 30 s, anelamento a 60 °C por 20 s e extensão a 72 °C por 1 min, seguidos de 10 min de extensão final a 72 °C. A reação continha 20 mM do par de oligonucleotídeos iniciadores, 3 mM de MgCl<sub>2</sub>, 0,8 mM dos 4 nucleotídeos e tampão de reação de PCR comercial (Invitrogen), em um volume final de 50 µL.

As amplificações foram confirmadas por eletroforese de DNA em gel de agarose 1%, em tampão TBE.

As reações foram purificadas para remoção dos iniciadores com 30% PEG/MgCl<sub>2</sub> (Polietileno glicol, Gateway™, Invitrogen). Após centrifugação, o sobrenadante foi cuidadosamente removido e descartado; o DNA precipitado foi submetido à secagem, por 5 minutos a 70 °C, e suspenso em 10 µL de água ultra-pura.

### 3.4.2 Vetor de entrada e vetor de expressão heteróloga

Os produtos de PCR purificados foram clonados no vetor de entrada pDONR™ 221 (Gateway™, Invitrogen). Para a reação, 50 fmol de inserto e 150 ng do

vetor foram recombinados pelo sistema sítio-específico, realizado pelo complexo enzimático comercial BP Clonase II, na presença de tampão TE pH 8,0 para um volume final de 10 µL. A reação aconteceu por 16 h a 25 °C e foi interrompida incubando-se 10 min a 37 °C com Proteinase K.

Os clones foram selecionados transformando-se 2 µL da reação em 50 µL de *Escherichia coli* DH5a cálcio-competente por choque térmico, de 1 min a 42 °C e 3 min em gelo. As células foram recuperadas por 1 h com 1 mL de meio LB a 37 °C, sob agitação de 200 rpm. 100 µL foram plaqueados em meio LB-ágar contendo canamicina para seleção do vetor e incubados a 37 °C por 16 h.

Os clones positivos foram determinados por PCR de colônia de bactérias, em uma reação semelhante à realizada para amplificar os genes de interesse, em um volume final de 10 µL. Os produtos foram analisados por eletroforese de DNA em gel de agarose.

As colônias selecionadas foram inoculadas em 10 mL de LB contendo canamicina a 100 µg/mL e incubadas a 37 °C, 200 rpm, por 16 h. Os plasmídeos foram extraídos com o kit MiniPrep (Qiagen) e quantificados no NanoDrop.

Os vetores foram submetidos à reação de recombinação com os vetores de destinação pDEST™ 17 (Gateway™, Invitrogen) para expressão heteróloga, e pTcGFP-C, pTcTAP-N e pTcPTP-C para transfecção em *T. cruzi* (BATISTA et al., 2010). A reação continha 150 ng de cada vetor (entrada e destinação) e o complexo enzimático comercial LR Clonase II (Gateway™, Invitrogen), em tampão TE pH 8,0 para um volume final de 10 µL.

Os clones positivos foram selecionados e extraídos conforme acima descrito.

### 3.4.3 Vetores de fusão para transfecção em *T. cruzi*

O gene codificante da proteína TcRBP40 foi anteriormente amplificado e clonado no vetor pDONR™ 221 da Plataforma Gateway(R), um sistema para recombinação sítio-específica. O gene foi recombinado nos vetores pTcPRGFPC e pTcPRGFNP de fusão carboxi- e amino-terminal, respectivamente, com a proteína GFP, para transfecção em *T. cruzi* (BATISTA et al., 2010). A reação de recombinação foi realizada de acordo com protocolo do fabricante do sistema Gateway. A reação foi

transformada em *E. coli* DH5 $\alpha$  para seleção de clones, e um deles escolhido para minipreparação de plasmídeo (QIAprep Spin Miniprep kit, Qiagen).

#### 3.4.4 Clonagem das regiões intergênicas no vetor repórter

A região intergênica à jusante do gene de TcRBP40 foi amplificada para clonagem no vetor pTCDUALuc (ARAÚJO et al., 2011). A sequência foi obtida do banco de dados TriTrypDB 4.0, a partir do cromossomo TcChr20-P. Foram desenhados iniciadores específicos, contendo os sítios de restrição *NheI* e *XhoI*, cujas sequências específicas foram: sentido TGCCACAATCGAAACCAAGAGC e anti-sentido GAATACAGGAAAGCAGACACAAGAATGC.

Para assegurar a digestão dos insertos com as enzimas de restrição, a região foi clonada no vetor pGEM-T® Easy (Promega), segundo especificações do fabricante, para então ser digerida com as enzimas específicas, bem como o vetor pTCDUALuc.

O inserto foi ligado no vetor pTcDUALuc em uma reação com a enzima T4 DNA Ligase (Invitrogen), segundo especificações do fabricante.

O vetor pTCDUALuc contendo a região intergênica foi transfectado em *T. cruzi* conforme descrito acima.

Os ensaios de localização dos transcritos dos genes repórter foram realizados conforme anteriormente publicado (ALVES et al., 2013).

### 3.5 Expressão de proteínas recombinantes heterólogas

Os vetores pDEST™ 17 contendo os 4 genes em estudo foram transformados na cepa de expressão *E. coli* BL21(DE3)pLysS cálcio-competentes (STUDIER et al., 1990), pelo método do choque térmico, conforme acima descrito.

3 a 4 colônias transformadas foram inoculadas em 25 ml de meio LB contendo ampicilina e cloranfenicol, resistências conferidas pelo vetor de expressão e pela cepa, respectivamente. Após crescimento a 37 °C em 200 rpm por cerca de 16 h, as culturas foram diluídas 1:20, num volume final de 500 mL de meio com antibióticos. Crescendo nas mesmas condições, após 2-3 h atingiram DO<sub>600</sub> entre 0,6-0,8. As culturas foram resfriadas até 15 °C e induzidas à expressão nesta temperatura com 0,1

mM de IPTG por cerca de 16 h, a 180 rpm. As células foram coletadas por centrifugação a 5.000 xg por 10 min, a 4 °C e ressuspendidas em tampão de lise solúvel, acrescido de coquetel inibidor de proteases (Roche). Para cada proteína foi padronizado um pH diferente de tampão, determinado de acordo com o ponto isoeletrico de cada proteína (ver Fig. x). Assim, TcRBP7 e TbRBP7A foram lisadas e purificadas em tampão Tris-HCl pH 9,5, enquanto TcRBP40 e TbRBP7B foram lisadas e purificadas em tampão fosfato de sódio pH 6,8.

A lise foi realizada por 6 pulsos de sonicação de 10 s, com intervalos de 1 min, a 80% da potência. O lisado foi centrifugado 20 min a 20.000 xg, a 4 °C, e o sobrenadante utilizado para a purificação das proteínas recombinantes ali presentes.

As frações foram analisadas por eletroforese em SDS-PAGE 15% e *Western blot*, para verificar a presença das proteínas expressas em cada fração.

### 3.5.1 Purificação de proteína recombinante solúvel

A expressão de um gene codificante pelo vetor pDEST™ 17 confere uma cauda N-terminal de 6 Histidinas à proteína expressa, permitindo sua captura por afinidade por uma resina mobilizada com Ni<sup>2+</sup>.

A fração solúvel (sobrenadante da lise) foi submetida a purificação por afinidade utilizando a resina Ni-NTA agarose (Qiagen). Assim, 5 mL do extrato solúvel foram incubados com 100 µL de resina, previamente equilibrada com o tampão da lise, por 1 h a 4 °C sob agitação leve. A resina foi filtrada em coluna de purificação e lavada 5 vezes com 2 mL do tampão de lise acrescido de concentrações crescentes de imidazol, para eliminar proteínas contaminantes fracamente ligadas: 2 vezes com 20 mM, 2 vezes com 50 mM e uma vez com 75 mM. As lavagens foram realizadas por 10 min sob agitação leve. A proteína recombinante foi eluída com 200 µL de tampão de lise acrescido de 250 mM de imidazol, em 5 frações.

As frações obtidas foram analisadas em eletroforese por SDS-PAGE 15%, para verificar a pureza das frações eluídas.

Alternativamente, as proteínas foram mantidas na resina, sem eluição, para prosseguir à captura *in vitro* de RNAs associados.

### 3.6 Ensaio molecular de captura de RNAs por *pull-down* com proteína recombinante

As proteínas TcRBP7 e TcRBP40 associadas à resina Ni-NTA Agarose (Qiagen) foram incubadas com 15-30 µg de RNA total de *T. cruzi*, extraído com o kit RNeasy Mini (Qiagen), na presença de 500 µL de tampão de ligação e inibidor de RNase, em triplicata biológica. Como competidores inespecíficos foram utilizados 10 µg de heparina e 5 mM de espermidina. Após 2 horas a 4 °C, sob agitação orbital, a fração não-ligada foi coletada por centrifugação de 1 min a 600 xg, 4 °C. A resina foi lavada 3 vezes com 1 mL de tampão de ligação, por 10 min a 4° C sob agitação leve.

Os RNAs foram extraídos da resina e da fração não-ligada com o kit RNeasy Mini.

O controle utilizado neste experimento foi a proteína GFP solúvel, contendo a mesma etiqueta de purificação, submetida ao ensaio de captura. A expressão e purificação da GFP obedeceu aos mesmos critérios utilizados para a produção da TcRBP7 e TcRBP40.

### 3.7 Sequenciamento de nova geração

Os RNAs capturados nos experimentos de ribonômica *in vivo* (imunoprecipitação) e *in vitro* (RNA *pull-down*) pelas proteínas TcRBP7 e TcRBP40 e GFP-controle foram identificados por sequenciamento em massa utilizando a plataforma SOLiD™ 4 System (Applied Biosystems).

Os RNAs foram preparados segundo as especificações do kit Total RNA-Seq (Applied Biosystems). Na etapa de purificação de fragmentos do gel de acrilamida, foram selecionados os fragmentos de cDNA entre 150 e 250 pb, já contendo os adaptadores.

Para cada amostra foi utilizado um *barcode* diferente, que permite a posterior identificação das amostras na população sequenciada. As bibliotecas continham um total de 20 amostras, em quantidades equivalentes. Foi preparada uma biblioteca para para cada lâmina de sequenciamento, sendo uma para a ribonômica *in vivo* e outra para a ribonômica *in vitro*, a fim de obter uma cobertura suficiente para cada amostra.

Para o experimento *in vivo*, tanto RNAs eluídos como os não-ligados das triplicatas biológicas foram sequenciados e analisados. Do experimento *in vitro* apenas os RNAs ligados foram sequenciados.

### 3.7.1 Análise do sequenciamento

O sequenciamento foi analisado utilizando o software CLC Genomics Workbench 5.5. O primeiro tratamento das leituras foi o *trimming*, onde cada fragmento sequenciado é processado para remoção das sequências adaptadoras presentes nas suas extremidades.

As leituras foram mapeadas frente ao genoma de *T. cruzi* CL Brener, haplótipo não-Esmeraldo, disponível no banco de dados TriTrypDB 8.0, acesso AAHK01. Foi determinado que as leituras apresentassem apenas 1 nucleotídeo de diferença com a sequência de referencia para o mapeamento ser válido, e que mapeassem em até 5 genes diferentes. Neste caso, as leituras mapeando em mais de um gene são proporcionalmente divididas entre eles de acordo com o número de leituras mapeadas unicamente para cada um dos genes. Isto aumenta a probabilidade da seleção dos genes corretamente capturados por cada RBP.

Um dos valores utilizados para a seleção dos alvos positivos de cada proteína foi o valor de RPKM (*reads/kilobase/million reads*) de cada amostra. O RPKM é um valor que representa o número de leituras de cada gene, dividido pelo tamanho do gene em kilobases, e dividido pelo total de leituras da amostra em milhões. O programa considera neste cálculo o total do número de leituras identificadas para o gene, e não somente o número de leituras unicamente mapeadas neste gene, mas obedecendo à proporção descrita no parágrafo anterior.

Este valor foi comparado entre a fração ligada e não-ligada de cada réplica, e foram considerados enriquecidos na fração ligada os genes apresentando uma média das razões (FC – *fold change*) maior ou igual a 2.

### 3.8 Análises computacionais

#### 3.8.1 Análise de motivos na região intergênica

A região intergênica à jusante dos genes codificantes de TcRBP40 e cruzipaína foi obtida a partir da base de dados TriTrypDB 4.0.

Os motivos em comum nas duas sequências foram preditos com o programa MEME 4.9.0 (*Multiple Em for Motif Elicitation*) (BAILEY; ELKAN, 1994).

#### 3.8.2 Análise filogenética

As sequências de DNA e de aminoácidos das proteínas TcRBP7, TcRBP40, TbRBP7A e TbRBP7B foram obtidas da base de dados TriTrypDB 1.1. Os números de acesso estão no quadro 2.

As análises de alinhamento dos aminoácidos, as construções das árvores e o teste de neutralidade Tajima D foram realizados com o software MEGA 4.0 (*Molecular Evolutionary Genetics Analysis*) (TAMURA et al., 2007). O alinhamento foi realizado com a ferramenta Clustal, a árvore filogenética de máxima verossimilhança pelo algoritmo *neighbor joining*, e a probabilidade calculada por 500 *bootstraps*.

**Quadro 2.** Genes empregados na análise filogenética da família RBP7 em tripanossomatídeos.

<b>Organismo</b>	<b>Acesso</b>	<b>Nome usual</b>
<i>L. braziliensis</i>	LbrM33_V2.1740	Lbra_RBP7B
<i>L. braziliensis</i>	LbrM33_V2.1720	Lbra_RBP7A
<i>L. braziliensis</i>	LbrM33_V2.1710	Lbra_RBP7C
<i>L. braziliensis</i>	LbrM26_V2.1550	Lbrazili_RBP8
<i>L. infantum</i>	LinJ33_V3.1540	Linf_RBP7A
<i>L. infantum</i>	LinJ33_V3.1560	Linf_RBP7B
<i>L. infantum</i>	LinJ33_V3.1530	Linf_RBP7C
<i>L. infantum</i>	LinJ26_V3.1510	Linfant_RBP8
<i>L. major</i>	LmjF33.1440	Lmajor_RBP7A
<i>L. major</i>	LmjF33.1450	Lmajor_RBP7B
<i>L. major</i>	LmjF33.1460	-
<i>L. major</i>	LmjF33.1470	Lmajor_RBP7D
<i>L. major</i>	LmjF26.1530	Lmajor_RBP8
<i>T. brucei</i>	Tb10.389.1650	Tbrucei_RBP7A
<i>T. brucei</i>	Tb10.389.1640	Tbrucei_RBP7B
<i>T. brucei</i>	Tb927.7.320	Tbrucei_RBP8
<i>T. brucei gambiense</i>	Tbgamb.32001	-
<i>T. brucei gambiense</i>	Tbgamb.32006	-
<i>T. congolense</i>	congo1497c05.q1k_2	Tcongo_RBP7
<i>T. cruzi</i>	TcCLB.506565.4	-
<i>T. cruzi</i>	TcCLB.506565.8	-
<i>T. cruzi</i>	TcCLB.506565.12	-
<i>T. cruzi</i>	TcCLB.511837.129	Tcruzi_RBP40
<i>T. cruzi</i>	TcCLB.511837.138	Tcruzi_RBP7
<i>T. cruzi</i>	TcCLB.508145.10	-
<i>T. cruzi</i>	TcCLB.508145.20	-
<i>T. cruzi</i>	TcCLB.508145.30	-
<i>T. cruzi</i>	TcCLB.504243.10	-
<i>T. cruzi</i>	TcCLB.504165.20	-
<i>T. cruzi</i>	TcCLB.508981.20	Tcruzi_RBP8
<i>T. vivax</i>	tviv1772d03.q1k_10	Tvivax_RBP7

### 3.8.3 Predição de estrutura tridimensional

A modelagem dos domínios RRM das quatro proteínas foi realizada em colaboração com o Dr. João Renato Carvalho Muniz, do Instituto de Física da USP. Os modelos foram construídos por modelagem estrutural por homologia, usando uma



estrutura do domínio de outro organismo já determinada por métodos experimentais (cristalografia e difração ou NMR), depositada no banco de dados PDB.

As estruturas obtidas foram visualizadas e as imagens tratadas no programa PyMOL (*The PyMOL Molecular Graphics System, Version 1.5.0.4* Schrödinger, LLC.).

Os cálculos de superfície eletrostática foram realizados com o *plugin* APBS 2.1 (*Adaptive Poisson-Boltzmann Solver*), no programa PyMOL.

## 4 RESULTADOS

### 4.1 Localização da proteína TcRBP40

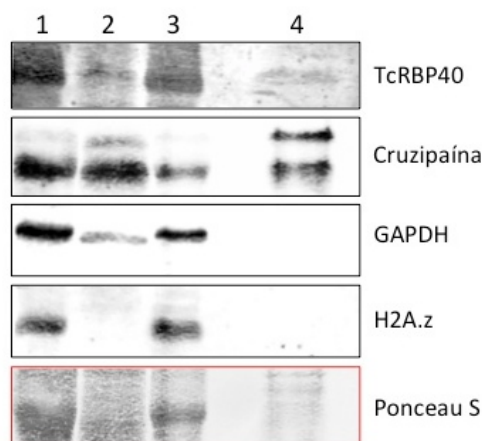
Dando sequência à caracterização funcional da proteína TcRBP40 – iniciada durante o mestrado anterior a esta tese – se faz importante determinar assertivamente sua localização celular, pois tal informação pode trazer pistas acerca da sua função biológica. Esta função vai além da propriedade de se associar a RNAs, conforme já demonstrado para esta proteína (GUERRA-SLOMPO et al., 2012) e inclui seu provável papel para a biologia do organismo.

#### *4.1.1 TcRBP40 está presente em uma fração enriquecida de reservossomos*

Em um trabalho anteriormente publicado, a proteína TcRBP40 foi reportada em foci citoplasmáticos específicos, na região posterior do parasita, co-localizando com a cruzipaína, uma proteína presente nos reservossomos do parasita (GUERRA-SLOMPO et al., 2012).

Uma abordagem diferenciada foi selecionada para determinar se a estrutura observada realmente se tratava de reservossomos. Foi utilizado um protocolo de fracionamento anteriormente estabelecido por Cunha-e-Silva et al. (2002), onde o lisado celular, sem a fração nuclear, é submetido a ultracentrifugação em um gradiente de sacarose. Desta forma, uma fração enriquecida de reservossomos foi obtida (poço 4) e as proteínas submetidas a identificação por *Western blot* (Fig. 6). Como controle, foram utilizados extratos contendo o lisado total de proteínas do parasita (poço 1), uma fração citoplasmática livre de organelas (poço 2) e uma fração pesada contendo organelas e o núcleo (poço 3). A proteína TcRBP40 foi identificada em todas as frações e nos reservossomos, juntamente com a cruzipaína, proteína de maior abundância no compartimento. O padrão de reconhecimento do anticorpo anti-cruzipaína utilizado (policlonal monoespecífico) mostra adicionalmente a forma não processada da cruzipaína, que tem maior massa molecular, e por isso são observadas duas bandas no resultado. A proteína H2A.z é uma histona, usada como marcador nuclear, enquanto GAPDH é uma proteína citoplasmática e também presente nos

glicossomos. A ausência destes marcadores na fração de reservossomos revela que a fração enriquecida não continha núcleo nem outras organelas co-purificadas.



**Figura 6.** A proteína TcRBP40 está presente em uma fração enriquecida de reservossomos. *Western blot* com diferentes extratos celulares: lisado total de proteínas de *T. cruzi* (1); fração citoplasmática livre de organelas (2); fração pesada contendo núcleo e organelas (3); fração enriquecida de reservossomos (4). Anti-TcRBP40 e anti-cruzipaína reconheceram as quatro frações. GAPDH aparece nas frações onde há organelas e também na citoplasmática, e H2A.z é vista apenas nas frações total e nuclear. A membrana foi corada com o corante reversível de proteínas Ponceau S, como controle da presença de proteínas em cada fração.

O soro policlonal contra TcRBP40 não mostrou uma boa especificidade no ensaio de *Western blot* contra o lisado total de proteínas, conforme já anteriormente exibido para os extratos do ciclo de vida do parasita (em GUERRA-SLOMPO et al., 2012). Isso pode ser explicado pela conservação do domínio RRM entre outras RBPs do mesmo organismo, ou ainda por outros epítomos que podem ser compartilhados com diversas proteínas.

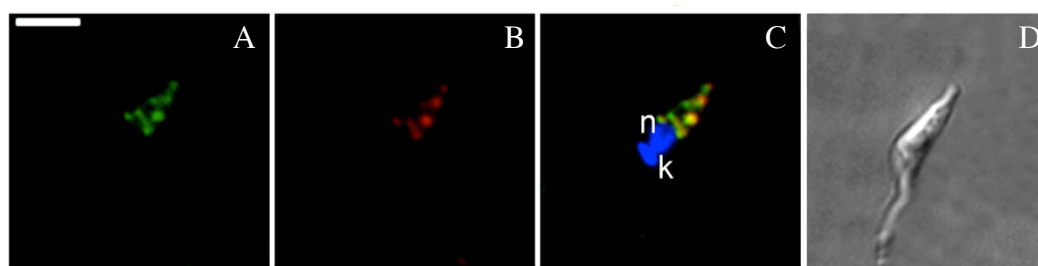
Por esta razão, ensaios complementares utilizando outras formas de localização foram realizados a fim de validar este resultado.

#### 4.1.2 A proteína TcRBP40 fusionada à GFP manteve a localização celular

Uma alternativa para realizar a localização celular de uma proteína independentemente de um anticorpo específico é fusionar a proteína a uma outra

molécula que permita sua localização. Assim, o gene codificante da proteína TcRBP40 foi clonado em um vetor de transfecção em *T. cruzi* para expressão episomal, o qual adiciona a proteína GFP à porção C-terminal da proteína TcRBP40 expressa. A fluorescência da GFP fusionada permitiu localizar a proteína TcRBP40 nos parasitas transfectantes, a qual apresentou o mesmo padrão observado pelo reconhecimento do anticorpo em experimento anterior (GUERRA-SLOMPO et al., 2012).

Utilizando a cultura transfectante da fusão TcRBP40-GFP foi realizado um ensaio de co-localização com o anticorpo contra a proteína cruzipaína. A sobreposição das imagens revelou o mesmo perfil de fluorescência, mostrando que a proteína fusionada apresenta a mesma localização da cruzipaína (Fig. 7).

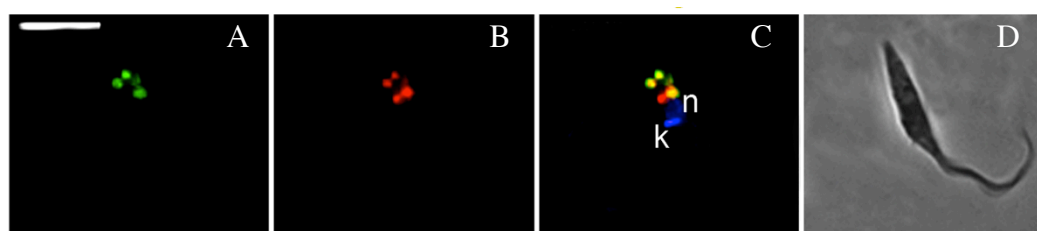


**Figura 7.** TcRBP40 fusionada a GFP co-localiza com cruzipaína. Imunofluorescência indireta de anti-cruzipaína (B) em parasita transfectado com vetor de fusão C-terminal da GFP à proteína TcRBP40 (A). Núcleo (n) e cinetoplasto (k) marcados com DAPI, no painel da sobreposição das fluorescências (C). D: contraste diferencial de interferência. Barra: 5  $\mu$ m.

#### 4.1.3 Os sítios da TcRBP40 são organelas funcionalmente ativas

Os reservossomos são estruturas de armazenamento nutricional e, portanto, o destino final da via endocítica, que se inicia na porção anterior do parasita, na região da bolsa flagelar e do citóstomo (SANT'ANNA; DE SOUZA; CUNHA-E-SILVA, 2004; SOARES; DE SOUZA, 1991). Explorando esta propriedade, é possível comprovar se os foci de fluorescência onde se observa a proteína TcRBP40 se tratam de reservossomos, observando a atividade endocítica da organela.

Para acompanhar o destino de partículas endocitadas, foi realizado um ensaio de incorporação de transferrina conjugada a Alexa 533. A localização da proteína TcRBP40 se deu pela fusão à GFP, cuja localização nos reservossomos foi demonstrada no ensaio anterior de co-localização com cruzipaína. Neste ensaio, os parasitas contendo a fusão TcRBP40-GFP foram submetidos a um curto estresse nutricional de 15 minutos, o qual não causa grandes impactos na biologia do parasita, mas o deixa mais propenso a englobar partículas que possa encontrar em seguida, no caso, a transferrina marcada com fluorescência (EGER; SOARES, 2012). As células ficaram expostas a esta única fonte nutricional por 30 minutos, tempo suficiente para que ocorra endocitose e destinação das vesículas endocitadas a compartimentos celulares, os reservossomos. Deste modo foi possível observar grandes vesículas contendo transferrina na porção posterior do parasita, e em várias delas, a presença da TcRBP40 fusionada a GFP (Fig. 8).



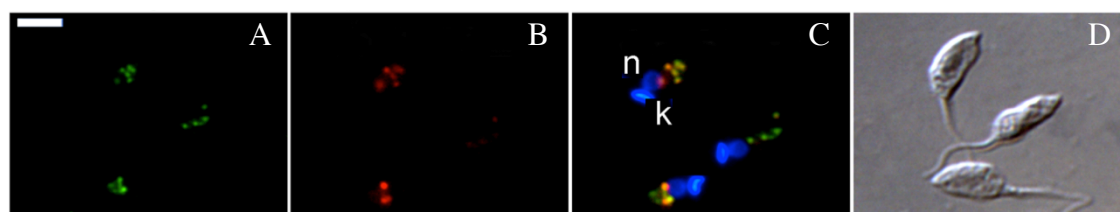
**Figura 8.** TcRBP40 se encontra em reservossomos funcionais. Parasita transfectante com a fusão TcRBP40-GFP (A) tratado com transferrina conjugada com Alexa 533 (B). Núcleo (n) e cinetoplasto (k) marcados com DAPI, no painel da sobreposição das fluorescências (C). As imagens foram tratadas em microscopia de deconvolução. À direita, contraste de fase (D). Barra: 5 µm.

#### 4.2 O mRNA da TcRBP40 se localiza nos reservossomos

Em um trabalho publicado recentemente foi mostrado que existe localização diferencial de mensageiros em *T. cruzi*, em combinação com a localização das proteínas por eles codificadas. O fenômeno foi observado nos mensageiros de proteína perinuclear (tubulina), flagelar (PFR2) e de reservossomos (cruzipaína) (ALVES et al., 2013).

Para localizar o mRNA codificante de TcRBP40 em formas epimastigotas de *T. cruzi* foi realizado um ensaio de hibridização *in situ* e microscopia de fluorescência (FISH). Para isso, uma sonda de DNA anti-sentido ao mRNA da TcRBP40 foi construída contendo na extremidade 5' a molécula fluorescente cianina-5 (Cy5). Como controle de especificidade da interação foi utilizada uma sonda sentido à sequência do mensageiro, contendo a mesma marcação.

Foi possível observar a localização do mRNA de TcRBP40 nas mesmas estruturas onde se encontra a proteína cruzipaina, conforme demonstrado pela co-localização da sonda com o anticorpo (Fig. 9).



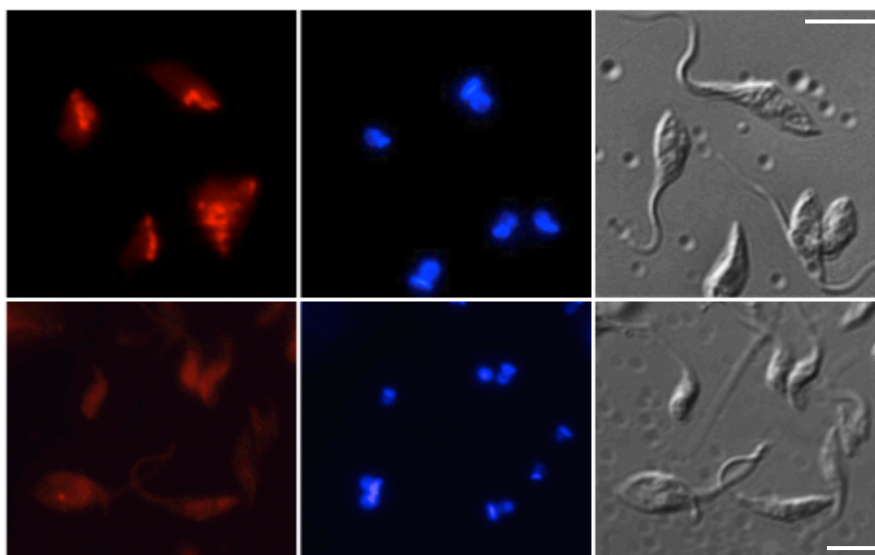
**Figura 9.** Localização do mRNA codificante da TcRBP40. FISH em epimastigota com sonda complementar ao mRNA de TcRBP40 conjugada com Cy5 na porção 5' (B), seguindo de imunofluorescência indireta com anti-cruzipaina e anti-IgG de camundongo conjugado a Alexa 488 (A). No painel C é possível constatar a sobreposição da marcação, na região posterior do parasita, pela sobreposição das imagens A e B. Núcleo (n) e cinetoplasto (k) marcados com DAPI. À direita, contraste diferencial de interferência (D). Barra: 5  $\mu$ m.

#### 4.2.1 O sinal de endereçamento do mRNA de TcRBP40 se encontra na região 3' UTR

Tendo sido demonstrada a localização celular especializada do mensageiro de TcRBP40 em *T. cruzi*, a questão seguinte foi determinar a região do transcrito onde se encontra um possível sinal de endereçamento para os reservossomos.

Um vetor desenvolvido por Araújo e colaboradores contém dois genes codificantes para a proteína luciferase, um proveniente da *Renilla* e outro do vagalume. O primeiro é regulado por regiões intergênicas de genes de expressão geral (controle), enquanto o segundo é regulado pela sequência nele inserida (teste) (ARAÚJO et al., 2011). O vetor foi gentilmente cedido pelo Dr. Wanderson D. daRocha, do Depto. de Bioquímica e Biologia Molecular, UFPR.

Os antecedentes com a cruzipáina mostraram que a região 3' UTR teria o potencial de endereçamento do mensageiro (ALVES et al., 2013). Assim, a região 3' intergênica da TcRBP40 foi clonada neste vetor, o qual foi transfectado em *T. cruzi*. O ensaio de localização por FISH utilizando sonda contra o mensageiro codificante da luciferase regulada mostrou a localização também na região posterior do parasita, como o perfil observado no FISH do mRNA de TcRBP40, enquanto a sonda contra o mensageiro da luciferase com regulador geral apresentou apenas um padrão disperso pelo citoplasma (Fig. 10). Isso indica que o possível elemento de endereçamento do mRNA de TcRBP40 está localizado na 3' UTR do transcrito.

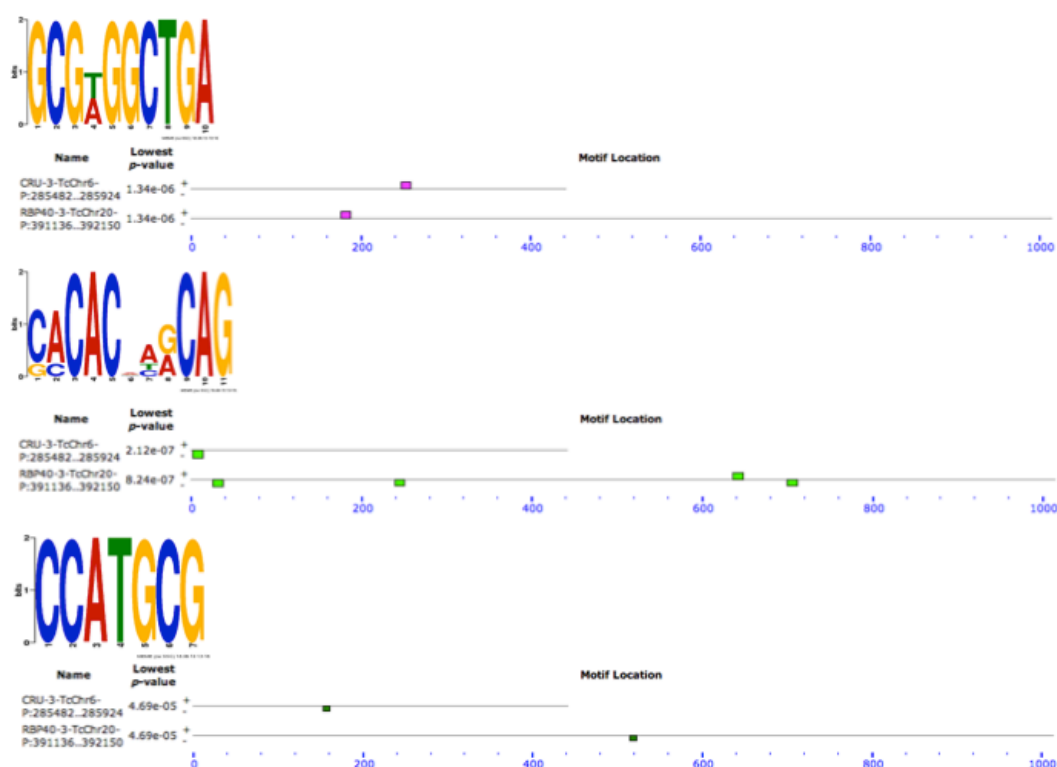


**Figura 10.** Expressão de gene repórter indicando localização definida por sinal de endereçamento da 3' UTR do mRNA de TcRBP40. FISH com sondas complementares ao mRNA de luciferase de *Renilla* ou de vagalume. O transcrito da luciferase de *Renilla* é flanqueado por regiões regulatórias de genes constitutivos, de expressão difusa. Já o da luciferase de vagalume é regulado pela porção 3' intergênica do gene de TcRBP40. No painel superior é possível observar uma localização predominante em pontos específicos na região posterior do parasita, enquanto no painel inferior a localização é mais dispersa pelo citoplasma. Núcleo (n) e cinetoplasto (k) marcados cm DAPI. À direita, contraste diferencial de interferência (DIC). Barra: 5  $\mu$ m.

#### 4.2.2 Os mRNAs de TcRBP40 e cruzipaína compartilham motivos em potencial nas suas regiões regulatórias

Os mRNAs codificantes de TcRBP40 e cruzipaína se localizam nos reservossomos. O ensaio de localização de um gene repórter mostrou que a região 3' UTR pode conter o sinal de endereçamento para esta organela em ambos os mensageiros (Fig. 10 e ALVES et al., 2013).

Assim, uma análise preliminar *in silico* buscou por elementos em comum entre as sequências intergênicas à jusante dos genes de TcRBP40 e cruzipaína. O programa MEME busca por motivos em comum entre as sequências ali inseridas. A figura 11 mostra 3 motivos semelhantes entre as duas sequências, posicionados dentro do que poderia ser considerado a 3' UTR dos mensageiros em análise.



**Figura 11.** Motivos conservados entre as regiões intergênicas à jusante de TcRBP40 e cruzipaína. Na representação há 3 motivos significativos e sua localização nas sequências, representados pelas caixas coloridas. Programa online MEME suite 4.9.1.



### 4.3 Análise filogenética

#### 4.3.1 Alinhamento da TcRBP40 contra proteínas de *T. brucei*: busca por homologia

A proteína TcRBP40 foi primeiramente descrita em Guerra-Slompo et al., 2012, como uma proteína da família *RNA Recognition Motif* (RRM). Formada por apenas 126 resíduos, é praticamente constituída pela região do domínio, delimitada por alguns aminoácidos na porção amino e uma pequena cauda carboxi-terminal.

O domínio RRM é estruturalmente conservado entre os eucariotos (CLÉRY; BLATTER; ALLAIN, 2008). Por esta razão, uma busca por homologia foi realizada, primeiramente buscando em um tripanossomatídeo próximo, o *T. brucei*.

Assim, a estrutura primária da TcRBP40 foi utilizada como sequência de entrada, para buscar na base de dados do *T. brucei* por proteínas contendo maior identidade à mesma. O alinhamento de sequências foi realizado utilizando o algoritmo BLAST, uma ferramenta de busca e alinhamento comparativo de sequências.

As primeiras sequências retornadas na busca foram das proteínas TbRBP7A e TbRBP7B, que contêm além de tamanho e características semelhantes à TcRBP40, uma alta taxa de identidade (74% e 68%) (Fig. 12). Isso pode ser um indicativo de proximidade de origem e função.

hypothetical protein [Trypanosoma brucei brucei strain 927/4 GUTat10.1]  
Sequence ID: [ref|XP\\_823414.1](#) Length: 116 Number of Matches: 1  
[▶ See 1 more title\(s\)](#)

Range 1: 3 to 113 [GenPept](#) [Graphics](#) ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
171 bits(433)	1e-55	Compositional matrix adjust.	82/111(74%)	89/111(80%)	0/111(0%)
Query 6	PRARLFVVGQLNFGATEDDIRAIFDFYGHVLSVNLHDKSKNRSKGSFAFVEYGSTGEADCA				65
Sbjct 3	PRARLFVVGQLNF ATEDDI A+FDYF L VN+L DK +SKGSFAFVEYGST EADCA				62
Query 66	IRSLHNRYMERDKPLQVSYCERSELISPFQWQALQLHQENSANPPPPDR				116
Sbjct 63	I +LHNRY MER+KPLQVSYC +SELIS FG A +LH EN ANP PP R				113

hypothetical protein [Trypanosoma brucei brucei strain 927/4 GUTat10.1]  
Sequence ID: [ref|XP\\_823415.1](#) Length: 116 Number of Matches: 1  
[▶ See 1 more title\(s\)](#)

Range 1: 3 to 113 [GenPept](#) [Graphics](#) ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
150 bits(378)	3e-47	Compositional matrix adjust.	76/111(68%)	81/111(72%)	0/111(0%)
Query 6	PRARLFVVGQLNFGATEDDIRAIFDFYGHVLSVNLHDKSKNRSKGSFAFVEYGSTGEADCA				65
Sbjct 3	PRARLFVVGQLNF ATEDDI A+FDYF VL VN+L DK +SKGSFAFVEYGST EAD A				62
Query 66	IRSLHNRYMERDKPLQVSYCERSELISPFQWQALQLHQENSANPPPPDR				116
Sbjct 63	I LHN R K LQVSYC +SELIS FG A +LH EN NP PP R				113

**Figura 12.** Alinhamento de sequências da proteína TcRBP40 como entrada, contra o banco de proteínas de *T. brucei*. No alinhamento superior está a proteína TbRBP7A como primeiro resultado, e no inferior a TbRBP7B. Ferramenta BLASTp online, em <https://blast.ncbi.nlm.nih.gov/Blast.cgi>.

#### 4.3.2 Alinhamento das proteínas TbRBP7A e B contra proteínas de *T. cruzi*

Para verificar se há homologia entre estas proteínas nos dois organismos, um alinhamento reverso foi realizado, nos mesmos parâmetros utilizados anteriormente. Nele, a sequência de entrada utilizada foi das proteínas TbRBP7A ou B, contra o banco de dados de proteínas de *T. cruzi*. A homologia seria confirmada ao retornar no alinhamento a TcRBP40 como melhor resultado.

No entanto, o primeiro resultado deste alinhamento foi a proteína TcRBP7, seguida da TcRBP40 em segundo lugar (Fig. 13). Isso indica que o homólogo das proteínas TbRBP7A e B de *T. brucei* é a proteína TcRBP7 de *T. cruzi*, e que TcRBP40 é uma proteína com alta taxa de identidade em relação às RBP7 dos dois organismos.

### RNA-binding protein [Trypanosoma cruzi strain CL Brener]

Sequence ID: [ref|XP\\_808057.1|](#) Length: 118 Number of Matches: 1

[► See 1 more title\(s\)](#)

Range 1: 5 to 115 [GenPept](#) [Graphics](#)

▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
167 bits(423)	8e-54	Compositional matrix adjust.	78/111(70%)	92/111(82%)	0/111(0%)
Query 4	RARLFVQQLNFNATEDDDIYALFDYGDVLHVNILRDKDTRKSKGSFVEYGSTVEADFAI				63
	+ARLFVQQLNF+ATE+D+ ALFDFYG VLHVN+LRD+ T +SKGSFVEYGST EAD AI				
Sbjct 5	QARLFVQQLNFDATEEDVCALFDYGVKVLHVNVLDRDTTNRKSKGSFVEYGSTEEADCAI				64
Query 64	LGLHNRNCTGHRKRLQVSYCNKSELISDFGRDHASKLHDENPVNPAPPQRV				114
	+ LHN R N K LQVSYC KSE+IS++GR+HA +LH EN NPAPP+R				
Sbjct 65	MALHNRNRYMERKPLQVSYCGKSEVISEYGREHALRLHRENSANPAPPPT				115

### RNA-binding protein [Trypanosoma cruzi strain CL Brener]

Sequence ID: [ref|XP\\_804683.1|](#) Length: 126 Number of Matches: 1

[► See 1 more title\(s\)](#)

Range 1: 6 to 116 [GenPept](#) [Graphics](#)

▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
150 bits(378)	6e-47	Compositional matrix adjust.	76/111(68%)	81/111(72%)	0/111(0%)
Query 3	PRARLFVQQLNFNATEDDDIYALFDYGDVLHVNILRDKDTRKSKGSFVEYGSTVEADFA				62
	PRARLFVQQLNF ATEDDI A+FDYF VL VN+L DK +SKGSFVEYGST EAD A				
Sbjct 6	PRARLFVQQLNFGATEDDIRAIFDFYGHVLSVNLHDKSKNRKSGSFVEYGSTGEADCA				65
Query 63	ILGLHNRNCTGHRKRLQVSYCNKSELISDFGRDHASKLHDENPVNPAPPQR				113
	I LHN R K LQVSYC +SELIS FG A +LH EN NP PP R				
Sbjct 66	IRSLHNRYYMERDKPLQVSYCERSELISPFQWQALQLHQENSANPPPPDR				116

**Figura 13.** Alinhamento de seqüências da proteína TbRBP7B como entrada contra o banco de proteínas de *T. cruzi*. O alinhamento superior é com a proteína TcRBP7 e o inferior com a TcRBP40. Ferramenta BLASTp.

O pareamento entre as proteínas TcRBP7 e TcRBP40 mostrou que elas de fato têm elevada identidade em sua seqüência primária. Estas análises podem indicar uma origem em comum destas RBPs.

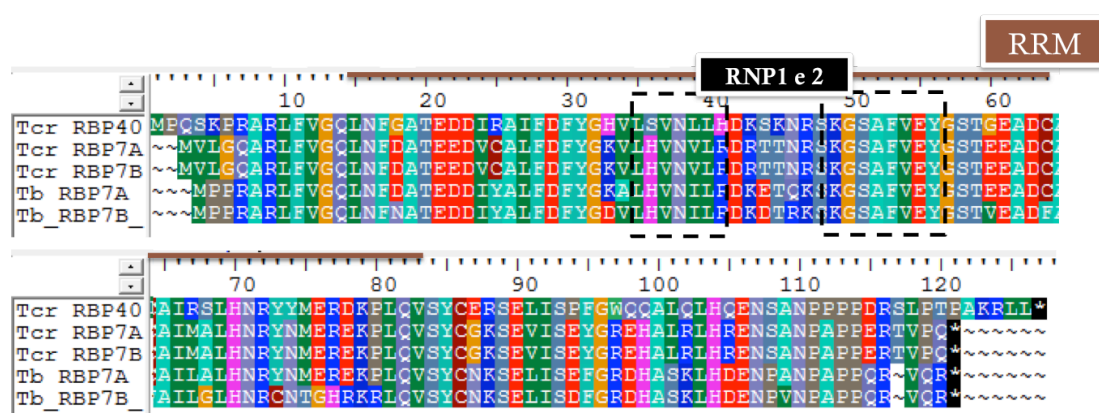
#### 4.3.3 Alinhamento das RBPs de *T. cruzi* e *T. brucei*

As RBPs em análise foram comparadas duas a duas, verificando a taxa comparativa nas seqüências de DNA e aminoácidos. A Tabela 1 mostra a porcentagem de identidade no alinhamento local entre os resíduos das RBP par a par.

**Tabela 1.** Comparação entre as quatro RBPs. Porcentagem de identidade entre as sequências protéicas. Análise por alinhamento de sequências local com a ferramenta BLASTp online.

	TcRBP7	TcRBP40	TbRBP7A
TcRBP40	72%		
TbRBP7A	81%	74%	
TbRBP7B	71%	68%	88%

Para visualizar as posições das divergências aminoacídicas foi realizado um alinhamento múltiplo global de sequências. Na figura 14 é possível observar que regiões variáveis se encontram no domínio RRM e na porção C-terminal das proteínas. A TcRBP40 apresenta diferenças mesmo dentro do motivo RNP2 de interação com RNA.



**Figura 14.** Alinhamento das proteínas TcRBP7 (duas cópias), TcRBP40, TbRBP7A e TbRBP7B com o programa BioEdit. A linha superior marrom indica a região do domínio RRM. As caixas pretas pontilhadas mostram a posição dos motivos RNP-1 e -2, os quais interagem com a molécula de RNA.

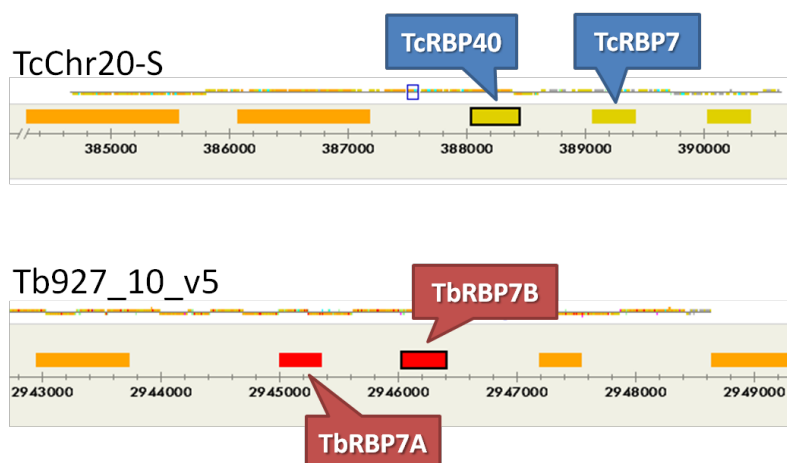
#### 4.3.4 Os genes homólogos de RBP7 estão adjacentes no genoma

c

Os genomas do *T. cruzi* e *T. brucei* são considerados compactos, ou seja, com regiões intergênicas reduzidas. Consequentemente, os genes adjacentes são próximos uns dos outros.

A análise do contexto da região cromossômica onde se inserem os genes de TcRBP7 e TbRBP7 utilizou informações das bases de dados GeneDB e TriTrypDB.

Intrigantemente, os genes TcRBP7 e TcRBP40 de *T. cruzi* e TbRBP7A e TbRBP7B de *T. brucei* são adjacentes no genoma dos respectivos organismos (Fig. 15). Isso seria indicativo de um evento de duplicação gênica dando origem às RBPs, as quais divergiram ao longo do tempo. Os demais genes da região permanecem sintênicos entre os organismos.



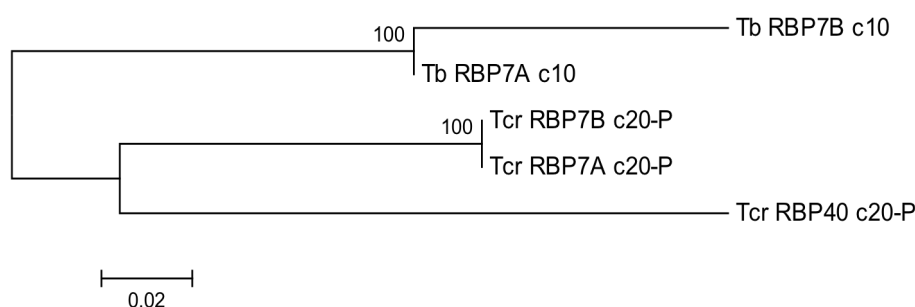
**Figura 15.** Posição cromossômica dos genes de RBP7 e RBP40 em *T. cruzi* e *T. brucei*. Ilustração de uma região dos cromossomos das cepas CL Brener de *T. cruzi* e 927 de *T. brucei brucei*. Adaptado da base de dados GeneDB.

Outro dado interessante é que podem existir outras cópias do gene TcRBP7 no genoma de *T. cruzi*, também em posição adjacente. O número de cópias não pôde ser determinado pelos dados genômicos e de alinhamento de cromossomos, pois as regiões intergênicas são também idênticas. De fato, devido a estas repetições, esta é uma região truncada no cromossomo. Somente uma abordagem por sequenciamento de regiões longas poderia determinar o real número de cópias de TcRBP7 no genoma.

#### 4.3.5 Os genes da família RBP7 em *T. cruzi* e *T. brucei* são duplicações gênicas

Estas quatro RBPs têm alta identidade entre elas, tamanho semelhante, domínio RRM, e mesma posição cromossômica.

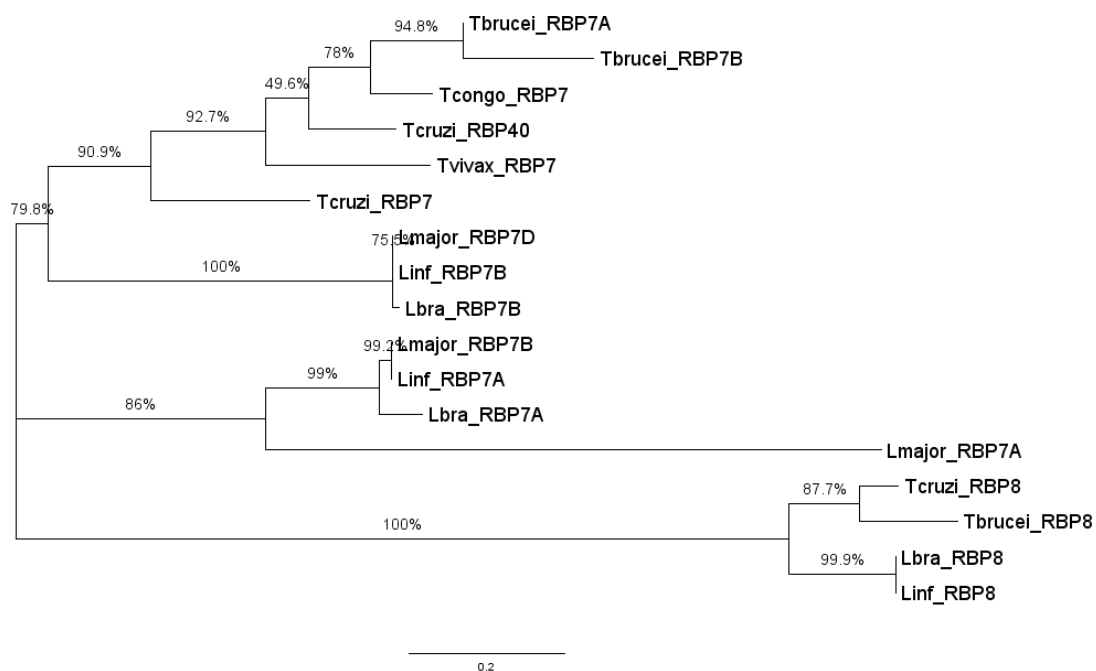
Uma vez que a evidência de duplicação e a alta semelhança possa indicar uma origem comum, foi realizada uma análise filogenética baseada no alinhamento das sequências de DNA dos quatro genes, para verificar a posição de cada um em relação a um provável ancestral comum. Na árvore filogenética, as RBPs se mostram divididas por organismo (Fig. 16). Isso é indicativo de que as duplicações gênicas aconteceram após a divergência entre as duas espécies, ou seja, ocorreu e se estabeleceu de modo independente e repetidamente nos dois organismos.



**Figura 16.** Árvore filogenética com os genes de TcRBP7 (mostrando duas das cópias) e TcRBP40 resultando no mesmo ramo, e TbRBP7A e TbRBP7B em outro ramo. Análise obtida a partir de alinhamento de sequências com a ferramenta Clustal, do programa MEGA 4.0, árvore de máxima verossimilhança, algoritmo *Neighbor-Joining*.

#### 4.3.6 O gene de RBP7 é também duplicado em outros tripanossomatídeos

Uma árvore filogenética foi construída a partir do alinhamento dos genes ortólogos de RBP7 de outros tripanossomatídeos (Fig. 17). É possível observar que os genes de *T. cruzi* e *T. brucei* estão agrupados, indicando serem eventos de duplicação exclusivos destes organismos, ou mais divergentes em relação ao ancestral. A tabela 2 mostra a distância entre os genes a partir do seu alinhamento.



**Figura 17.** Árvore filogenética dos ortólogos de RBP7 em tripanossomatídeos, construída a partir de alinhamento múltiplo de sequências de aminoácidos. A proteína RBP8 foi utilizada como grupo externo na análise. Programa MEGA 4.0, máxima verossimilhança, algoritmo *neighbor-joining*. Valores representam probabilidades *bootstrap*. Análise com 1000 *bootstraps*.

**Tabela 2.** Matriz de distâncias entre os genes de RBPs em tripanossomatídeos. Os valores correspondem ao número de substituições de base por sítio entre cada par de sequências. Programa MEGA, modelo MCL.

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	Tcruzi_RBP7																
2	Tcruzi_RBP40	0,41															
3	Tbrucei_RBP7A	0,46	0,33														
4	Tbrucei_RBP7B	0,48	0,41	0,13													
5	Lmajor_RBP7A	0,84	0,76	0,81	0,82												
6	Lmajor_RBP7B	0,6	0,55	0,67	0,72	0,59											
7	Lmajor_RBP7D	0,62	0,54	0,61	0,65	0,76	0,54										
8	Lbra_RBP7B	0,64	0,51	0,61	0,67	0,75	0,56	0,08									
9	Linf_RBP7A	0,59	0,53	0,65	0,7	0,6	0,01	0,55	0,56								
10	Linf_RBP7B	0,63	0,55	0,61	0,66	0,75	0,55	0,04	0,07	0,56							
11	Lbra_RBP7A	0,62	0,58	0,68	0,72	0,6	0,12	0,61	0,6	0,12	0,62						
12	Tvivax_RBP7	0,52	0,38	0,4	0,49	0,85	0,64	0,57	0,57	0,62	0,58	0,67					
13	Tcongo_RBP7	0,42	0,31	0,28	0,39	0,8	0,63	0,62	0,59	0,62	0,61	0,66	0,36				
14	Tcruzi_RBP8	0,82	0,93	0,86	0,92	1,03	0,86	0,96	0,95	0,86	0,97	0,83	0,97	0,82			
15	Tbrucei_RBP8	0,89	0,78	0,75	0,77	0,89	0,78	0,83	0,87	0,77	0,84	0,77	0,83	0,75	0,34		
16	Lbra_RBP8	0,97	0,86	0,98	1,05	0,91	0,75	0,82	0,82	0,73	0,81	0,76	0,95	0,95	0,41	0,44	
17	Linf_RBP8	0,94	0,92	0,98	0,99	0,88	0,7	0,83	0,82	0,68	0,83	0,71	0,98	0,9	0,41	0,43	0,09



#### 4.3.7 Os genes de *TbRBP7A* e *TbRBP7B* acumulam mutações assimetricamente

TcRBP7 e TcRBP40 apresentam diferenças aminoacídicas em diversas regiões, inclusive no sítio de interação com o RNA. Já as proteínas TbRBP7A e TbRBP7B são muito semelhantes, e seu DNA apresenta 88% de identidade.

Para avaliar a neutralidade das divergências no DNA entre os genes de *T. brucei*, foi empregado o teste estatístico Tajima D. Neste teste, as posições dos nucleotídeos são comparadas nos dois genes. A comparação foi realizada em relação a um terceiro gene, o ortólogo de RBP7 em *T. congolense*, escolhido por ser o mais próximo entre os ortólogos dos outros organismos (ver Fig. 17).

A Tabela 3 mostra que a posição de maior variabilidade é a 3ª no quadro de leitura do ribossomo, que é majoritariamente neutra em relação a trocas aminoacídicas. O dado mais significativo é sobre a maior representatividade de bases exclusivamente encontradas em TbRBP7B em todos os testes (colunas), o que indica que há mais mutações acumuladas neste gene do que em TbRBP7A em relação ao grupo externo. Isso significa que os genes duplicados estão se distanciando e que TbRBP7B está divergindo mais em relação a um ancestral comum e, portanto, acumulando mutações mais rapidamente.

**Tabela 3.** Análise da divergência de bases nos genes de *T. brucei*. Teste Tajima.

Contagem de bases	1ª base	2ª base	3ª base	1ª e 2ª bases	1ª, 2ª e 3ª bases	AAs
Idênticas nas 3 RBPs	100	94	38	194	232	84
Divergentes nas 3 RBPs	0	0	7	0	7	3
Exclusivas em TbRBP7A	0	0	6	0	6	0
Exclusivas em TbRBP7B	3	12	12	15	27	13
Exclusivas em TcongoRBP7	12	9	52	21	73	15
d	3	12	2	15	13,36	13
p-value*	8,33E-02	5,30E-04	1,57E-01	1,10E-04	2,60E-04	3,10E-04

\*O valor de p menor do que 0,05 é usado para rejeitar a hipótese nula de taxa igual entre os genes.

d: valor estatístico.

AAs: aminoácidos.

#### 4.4 Comparação molecular entre as RBPs

As proteínas TcRBP7, TcRBP40, TbRBP7A e TbRBP7B foram comparadas segundo as suas propriedades enquanto proteínas. O Quadro 3 informa características elementares como tamanho, massa e ponto isoelétrico, e também índices sobre a estabilidade estrutural das proteínas. Brevemente, o programa descreve o coeficiente de extinção como um indicativo de absorbância de luz pela proteína em certo comprimento de onda. O índice de instabilidade é um cálculo desta propriedade em um tubo de teste. O índice alifático é definido como o volume relativo ocupado por cadeias laterais alifáticas, sendo positivo com o aumento da termoestabilidade de proteínas globulares. E o GRAVY é calculado pela soma dos valores de hidropatias de todos os aminoácidos, dividido pelo número de resíduos.

**Quadro 3.** Comparação molecular entre as 4 proteínas. Valores obtidos pelo programa de análises ProtParam, disponível no portal Expasy de análises de proteínas. O gradiente de cores indica os valores maiores, intermediários e menores dentro do grupo, iniciando pelo verde até o vermelho, conforme a barra indicadora. pI: ponto isoelétrico.

	TcRBP7	TcRBP40	TbRBP7A	TbRBP7B
Resíduos	118	126	116	116
Massa (kDa)	14,32	13,41	13,27	13,22
pI teórico	5,79	8,62	5,78	7,84
Resíduos negativos	17	14	18	15
Resíduos positivos	14	16	15	16
Coeficiente de extinção	7450	12950	7450	5960
Índice de instabilidade	43,15	58,46	45,87	40,11
Índice alifático	73,56	71,27	69,05	73,97
GRAVY	-0,536	-0,662	-0,742	-0,64

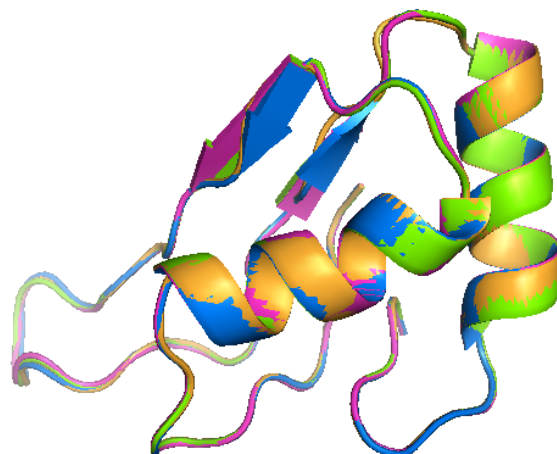


De um modo geral, as proteínas de *T. cruzi* apresentaram os maiores índices, e as de *T. brucei* os menores. TcRBP7 se aproximou mais de TbRBP7A. Intrigantemente, as proteínas TbRBP7A e TbRBP7B, apesar de conservarem alta identidade de sequência, tiveram índices menos semelhantes entre si do que entre TbRBP7A e TcRBP7. Esta análise demonstra que há divergência nas propriedades das proteínas duplicadas em cada organismo.

#### 4.4.1 Modelagem molecular

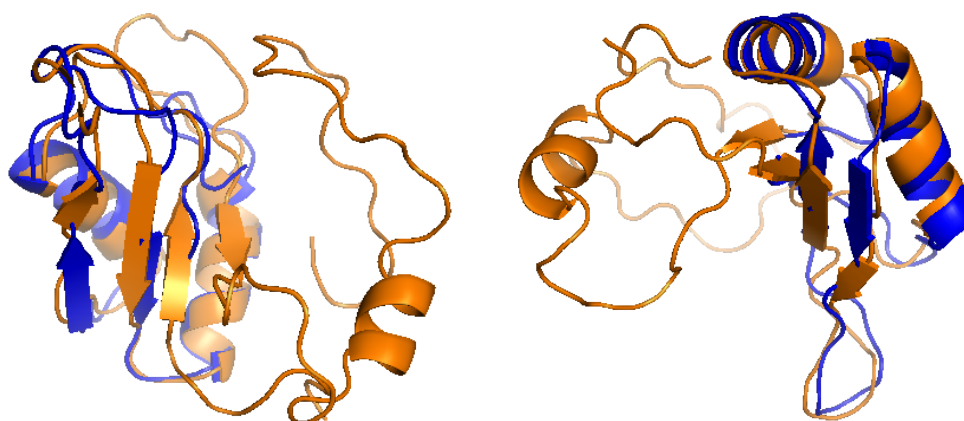
A modelagem molecular por homologia é a montagem tridimensional das estruturas secundária e terciária de uma proteína, baseada na homologia da sua sequência primária (aminoacídica) com outra proteína, com a qual compartilha alta identidade, que tenha sua estrutura resolvida por métodos experimentais (cristalografia, NMR – Ressonância Magnética Nuclear).

Através de uma colaboração com o Dr. João Renato Muniz, do Instituto de Física da Universidade de São Paulo (USP), as proteínas TcRBP7, TcRBP40, TbRBP7A e TbRBP7B foram analisadas quanto à sua estrutura tridimensional putativa. Elas foram construídas baseando-se no modelo conhecido de outra RBP, depositado no banco de dados de proteínas PDB. No entanto, apenas o domínio RRM das proteínas em estudo foi utilizado nesta análise, o qual apresenta uma estrutura muito conservada entre os eucariotos. As imagens foram geradas pelo programa PyMol. Na Figura 18 é possível observar a sobreposição total do domínio nas quatro RBPs. Isto acontece porque a modelagem é feita utilizando-se uma estrutura como “molde”, a qual tem uma razoável identidade de sequência com a proteína em estudo. Por esta razão as regiões externas ao domínio RRM não puderam ser modeladas nesta metodologia, já que são sequências muito variáveis entre os organismos e os programas poderiam assumir qualquer estrutura desejada, baseado na rotação de cada resíduo.



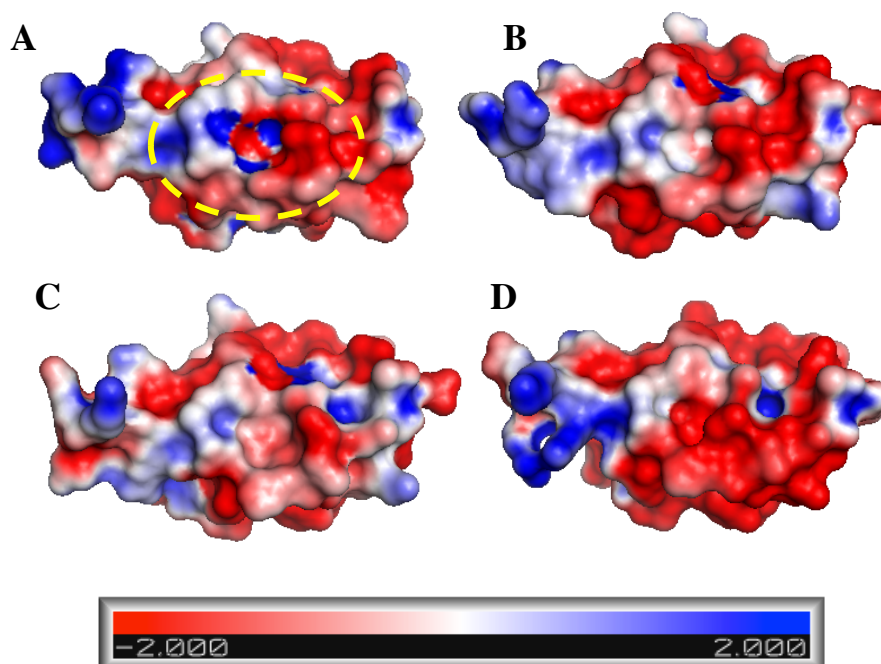
**Figura 18.** Modelo do domínio RRM das quatro RBPs. Representação em desenho dos domínios de cada proteína sobrepostos, demonstrados pelas diferentes cores. Na representação são vistas apenas 2 fitas  $\beta$  ao invés de 4, devido ao caráter parcial da sequência utilizada para a modelagem.

Utilizando outro algoritmo de modelagem molecular, a proteína TcRBP40 teve sua estrutura completa modelada, incluindo as regiões externas ao domínio RRM (Fig. 19). É possível observar que o restante da proteína não apresenta uma estrutura bem determinada, formada principalmente por *loops* desconexos. As regiões N- e C-terminal da TcRBP40 poderiam assumir qualquer conformação possível, desde formar um sítio acessório ao RRM até envolver parte do domínio e manter algumas superfícies encobertas deste. Fica assim evidente o desafio enfrentado ao modelar estruturas de proteínas *de novo*, sem o auxílio de uma estrutura molecular já conhecida. Apesar disso, ambos os modelos podem ser sobrepostos em grande parte da molécula.



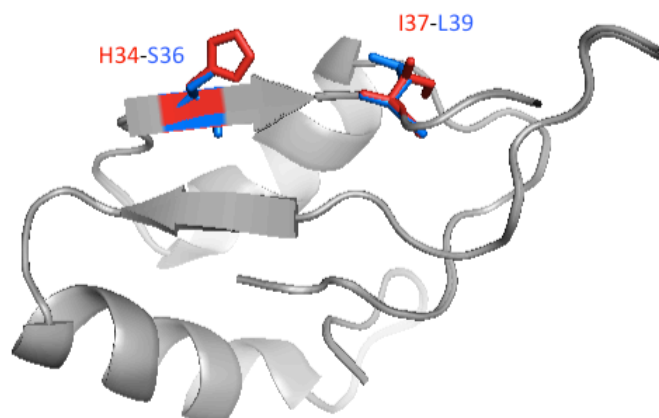
**Figura 19.** Alinhamento tridimensional da TcRBP40. Representação em desenho da sobreposição entre a estrutura parcial, contendo apenas o domínio RRM (azul), e a proteína inteira (laranja), vistas em dois ângulos diferentes.

Analisando outras informações dos modelos construídos, as proteínas TcRBP7, TcRBP40, TbRBP7A e TbRBP7B foram comparadas quanto à sua superfície eletrostática (Fig. 20). As diferenças são sutis, mas podem ser consideráveis comparando-se o sítio de interação com o RNA, formado pelas duas fitas  $\beta$ , ao centro das imagens. Entre as proteínas de *T. cruzi* (Figs. 20A e B) há pouca diferença nesta região do centro da proteína, excetuando-se a porção côncava central de cores mescladas em TcRBP40, que constitui um erro do reconhecimento das cargas destes aminoácidos. Já as extremidades à direita e esquerda são regiões mais variáveis, pois são formadas por resíduos que não estão interagindo com o *core* central do domínio. As proteínas TcRBP7 e TbRBP7A (Figs. 20B e C) apresentaram maior semelhança comparativa deste sítio, enquanto a proteína TbRBP7B (D) foi a mais divergente dentre as quatro do grupo, tanto em superfície quanto em distribuição de cargas.



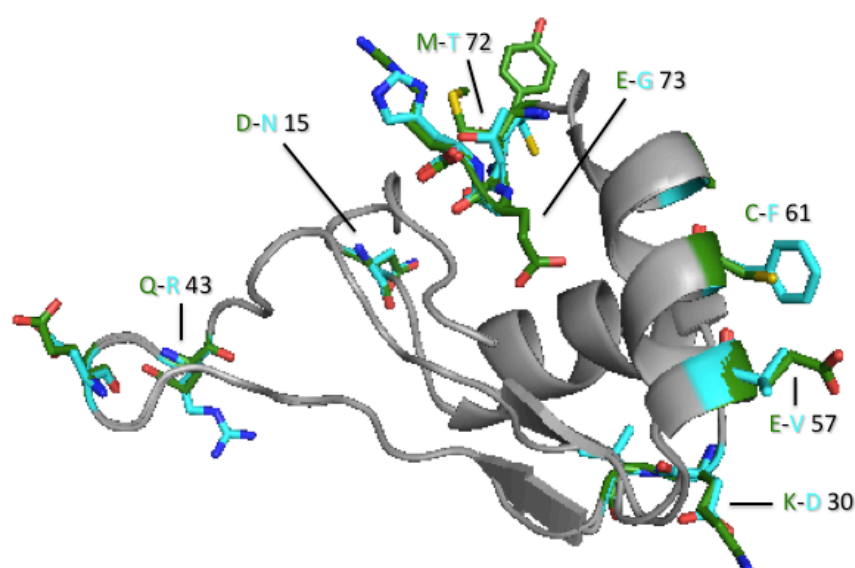
**Figura 20.** Mapa eletrostático de superfície das 4 RBPs. O ângulo de visão selecionado para as quatro proteínas mostra as duas fitas  $\beta$ , que são os dois motivos de interação direta com o RNA RNP-1 e RNP-2, indicado com um círculo amarelo pontilhado em A. A) TcRBP40, B) TcRBP7, C) TbRBP7A, D) TbRBP7B. O código de cores indica a carga elétrica da região: negativo em vermelho e positivo em azul. Os valores da escala são o parâmetro de escolha da intensidade para representação das cargas, igual para as 4 proteínas.

Os aminoácidos divergentes foram comparados em duplas de proteínas. TcRBP7 e TcRBP40, as duas proteínas de *T. cruzi*, apresentam divergência em dois resíduos dentro do motivo RNP-2 de interação com o RNA (Fig. 21), o que pode determinar diversidade na interação com os alvos. A diferença entre His e Ser é importante, pois se tratam de aminoácidos pertencentes a grupos distintos, e em uma região crítica para o reconhecimento do elemento em *cis* no RNA. Assim, mesmo tendo alta identidade em estrutura primária e secundária, estas RBPs podem se associar a alvos distintos (especificidade) ou com intensidade diferente (afinidade).



**Figura 21.** Alinhamento estrutural entre as proteínas TcRBP7 e TcRBP40. Comparação da estrutura do domínio RRM de ambas as proteínas quanto a diferenças residuais no motivo RNP-2, localizado em uma fita  $\beta$ , onde são destacadas duas divergências: entre Histidina e Serina e entre Isoleucina e Leucina, onde vermelho representa os resíduos da TcRBP7 e azul da TcRBP40. Os números indicam a posição do resíduo em cada proteína.

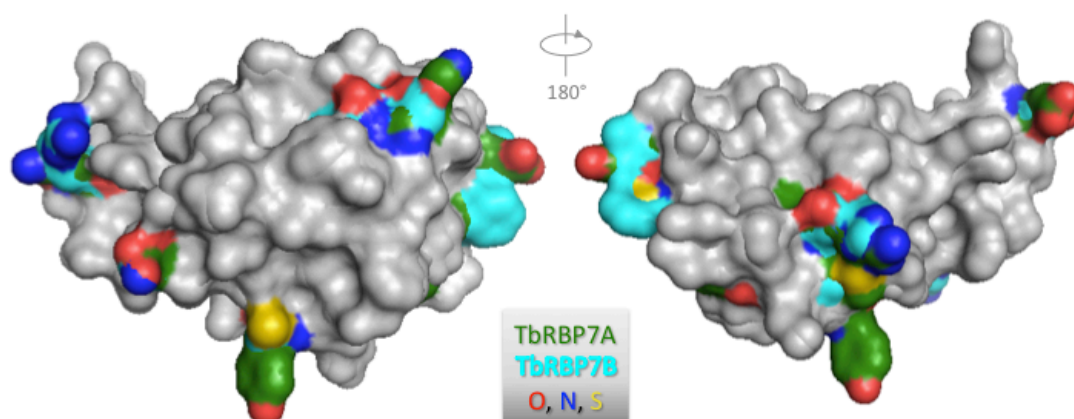
Na comparação das duas proteínas de *T. brucei*, as quais compartilham 88% de identidade, foram evidenciadas as suas diferenças residuais. Nenhuma delas está presente nos motivos RNP-1 e RNP-2, mas em outras regiões que podem fazer parte da estrutura externa da proteína. A Figura 22 mostra todas as posições de divergência, e foram nomeados os resíduos pertencentes a grupos de aminoácidos diferentes.



**Figura 22.** Alinhamento estrutural entre os RRM de TbRBP7A e TbRBP7B. Desenho da sobreposição dos domínios das duas proteínas de *T. brucei*. As cadeias laterais dos resíduos divergentes são exibidas para comparação estrutural. Os resíduos

pertencentes a grupos de aminoácidos distintos são mostrados juntamente com sua posição na sequência primária das proteínas. Em verde os resíduos da TbRBP7A e em azul ciano da TbRBP7B, enquanto em cinza estão os resíduos não variáveis, sem suas cadeias laterais. As cores variadas nas extremidades das cadeias laterais indicam diferentes átomos: azul royal para Nitrogênio, vermelho para Oxigênio e amarelo para Enxofre.

Observando a superfície externa do domínio nas proteínas TbRBP7A e TbRBP7B é possível identificar a posição das divergências dos aminoácidos (Fig. 23). Não há diferenças significativas na superfície da proteína, comparando-se as cores em cada modelo, a não ser por algumas extremidades modificadas, expondo diferentes átomos a interações com outras moléculas na célula.

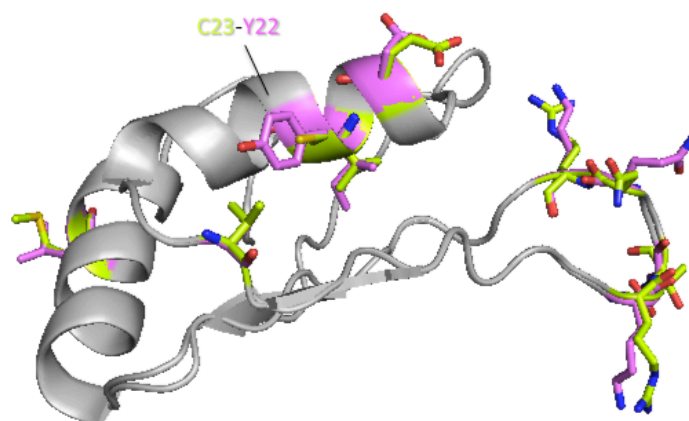


**Figura 23.** Sobreposição das superfícies do domínio RRM de TbRBP7A e TbRBP7B. Cada imagem mostra a sobreposição entre os domínios em rotações diferentes. As regiões em cinza são regiões iguais entre as proteínas, enquanto as regiões em verde são exclusivas da TbRBP7A e em azul ciano da TbRBP7B. As extremidades em vermelho representam átomos de Oxigênio (O) dos resíduos divergentes, em azul royal de Nitrogênio (N) e em amarelo de Enxofre (S).

As proteínas TcRBP7 de *T. cruzi* e TbRBP7A de *T. brucei* são as que apresentam maiores semelhanças moleculares entre as duas espécies, conforme demonstrado pelo Quadro 3. Então o alinhamento estrutural foi realizado a fim de comparar os domínios RRM destas duas proteínas. A Figura 24 mostra a sobreposição estrutural, onde é possível observar apenas uma divergência de resíduos pertencentes a grupos diferentes. É, ainda assim, uma divergência importante, pois o resíduo Cisteína é capaz de realizar pontes dissulfeto com outras proteínas, o que confere um sítio com potencial para divergência de interações comparando-se funcionalmente as duas

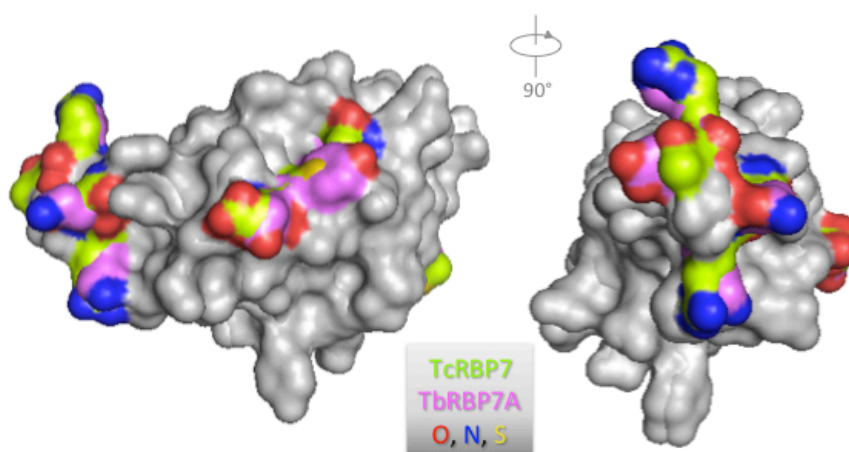


proteínas. Os sítios de interação com o RNA são idênticos em sequência (fitas  $\beta$ , voltadas para baixo).



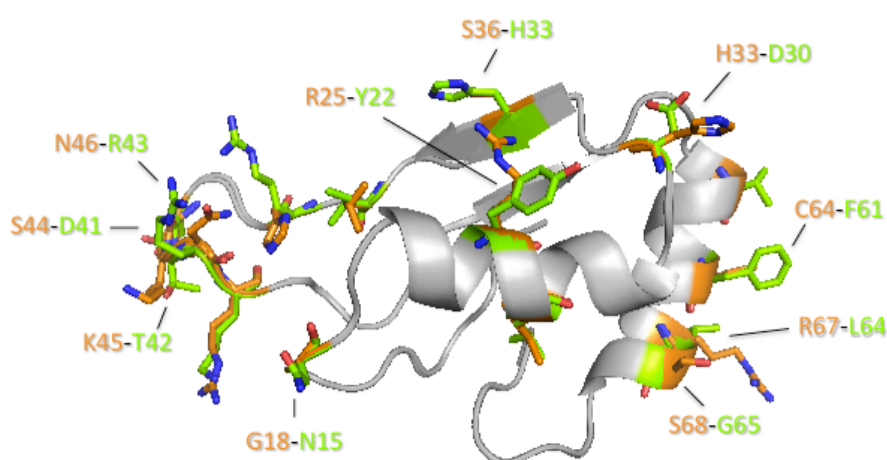
**Figura 24.** Alinhamento estrutural dos domínios RRM das proteínas TcRBP7 e TbRBP7A. Desenho esquemático das regiões do domínio RRM nas duas proteínas, mostrando as cadeias laterais apenas dos resíduos divergentes. As regiões em cinza são idênticas em sequência. Regiões em verde representam os resíduos da TcRBP7 e em rosa da TbRBP7A. As letras indicam o código dos resíduos divergentes pertencentes a grupos diferentes, e os números são a posição dos resíduos na respectiva proteína, obedecendo ao mesmo código de cores. As cores nas extremidades das cadeias laterais coloridas indicam diferentes átomos: azul royal para Nitrogênio, vermelho para Oxigênio e amarelo para Enxofre.

A superfície da comparação estrutural entre os domínios de TcRBP7 e TbRBP7A mostra a posição das divergências residuais. Apenas duas regiões são divergentes entre as duas superfícies (Fig. 25), podendo indicar uma alta conservação funcional entre estas proteínas representadas.



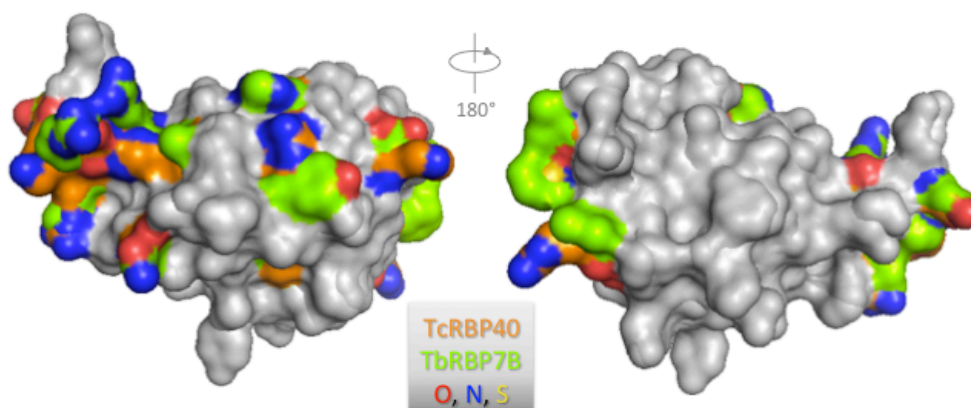
**Figura 25.** Sobreposição das superfícies do domínio RRM de TcRBP7 e TbRBP7A. As imagens mostram a sobreposição das estruturas, em ângulos rotacionais diferentes. As regiões em cinza são idênticas em sequência. Regiões em verde representam os resíduos divergentes da TcRBP7 e em rosa da TbRBP7A. As cores nas extremidades das cadeias laterais indicam diferentes átomos: azul royal para Nitrogênio, vermelho para Oxigênio e amarelo para Enxofre.

O alinhamento da estrutura do domínio RRM de TcRBP40 de *T. cruzi* e TbRBP7B de *T. brucei*, as mais diferentes molecularmente dentre as RBPs em estudo, mostrou que há uma série de divergências não-sinônimas entre os resíduos (Fig. 26).



**Figura 26.** Sobreposição estrutural do domínio RRM das proteínas TcRBP40 e TbRBP7B. Desenho esquemático das regiões do domínio RRM nas duas proteínas, mostrando as cadeias laterais apenas dos resíduos divergentes. As regiões em cinza são idênticas em sequência. Regiões em laranja representam os resíduos da TcRBP40 e em verde da TbRBP7B. As letras indicam apenas os resíduos divergentes pertencentes a grupos diferentes de aminoácidos, e os números são a posição dos resíduos nas respectivas proteínas, obedecendo ao mesmo código de cores. As cores nas extremidades das cadeias laterais indicam diferentes átomos: azul royal para Nitrogênio, vermelho para Oxigênio e amarelo para Enxofre.

A posição destas divergências na superfície dos domínios mostrou vários sítios afetados (Fig. 27). Ainda assim, uma grande região se manteve conservada entre os dois domínios, a qual é correspondente às fitas  $\beta$ , sítio dos motivos RNP-1 e RNP-2.



**Figura 27.** Sobreposição das superfícies do domínio RRM de TcRBP40 e TbRBP7B. As duas imagens mostram a sobreposição da superfície dos domínios, em ângulos rotacionais diferentes. As regiões em cinza são idênticas em sequência. Regiões em laranja representam os resíduos divergentes da TcRBP40 e em verde da TbRBP7B. As cores nas extremidades das cadeias laterais indicam diferentes átomos: azul royal para Nitrogênio, vermelho para Oxigênio e amarelo para Enxofre.

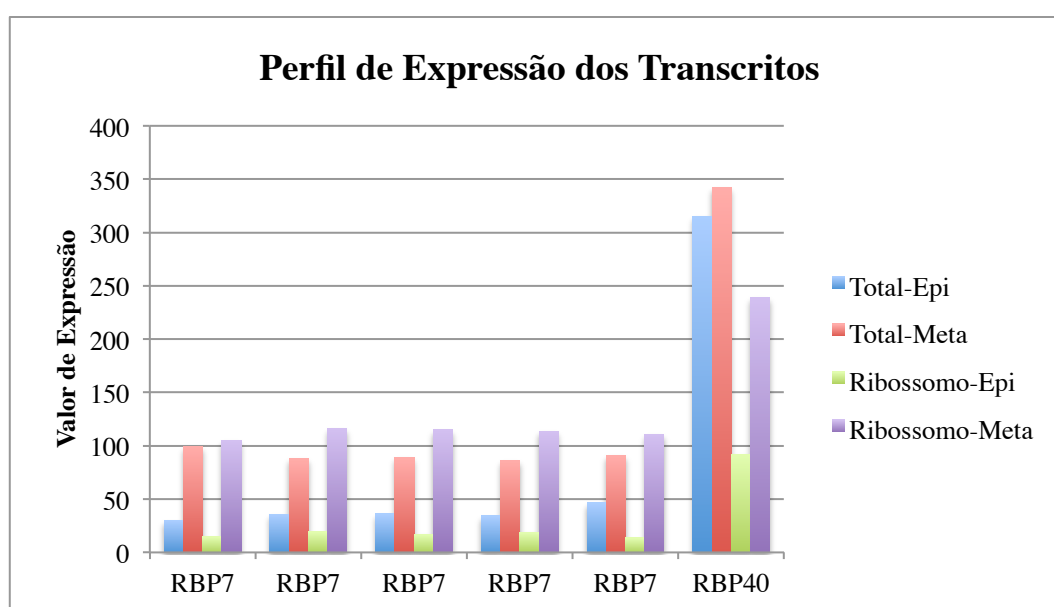
Todas estas divergências estruturais observadas nas 4 RBPs podem refletir diversificação funcional entre as proteínas, tanto em relação aos RNAs alvos por elas regulados, quanto em relação aos parceiros aos quais se associam para a regulação, embora esta diversificação não possa ser qualificada por este tipo de análise. Após este trabalho estrutural, a caracterização funcional das proteínas pode indicar o quanto cada conservação ou divergência foi importante para o que se venha a observar da função de cada RBP.

#### 4.4 Caracterização funcional das proteínas TcRBP7 e TcRBP40

##### 4.5.1 Os genes de TcRBP7 e TcRBP40 apresentam perfis proporcionais de expressão em epimastigotas e tripomastigotas metacíclicas

Os transcritos de TcRBP7 e TcRBP40 foram analisados experimentalmente em um estudo em larga-escala sobre a mobilização polissomal dos genes de *T. cruzi*, como parte da tese da Dra. Saloê Bispo Poubel (POUBEL, 2014). Neste estudo foram analisados os transcritomas das formas epimastigotas e tripomastigotas metacíclicas por RNA-Seq. Tanto os mRNAs totais da célula quanto os mRNAs associados aos ribossomos foram analisados qualitativa e quantitativamente.

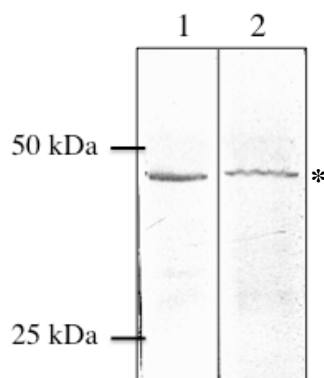
A Figura 28 mostra que os transcritos de TcRBP7 e TcRBP40 têm perfis de expressão proporcionalmente semelhantes, tanto nos mRNAs totais da célula quanto associados a ribossomos. No gráfico são mostradas cinco cópias do gene de TcRBP7, todas apresentando perfil de expressão semelhante. É possível que os valores unidos de cada gene de TcRBP7 sejam aproximados aos valores apresentados para TcRBP40, uma vez que as 5 cópias codificam um mesmo transcrito, resultando em proteínas idênticas.



**Figura 28.** Perfil de expressão dos genes de TcRBP7 e TcRBP40. Dados de RNA-Seq de mRNAs totais de formas epimastigotas (Epi) ou tripomastigotas metacíclicas (Meta) e de mRNAs associados aos ribossomos nas duas formas do ciclo de vida. Os valores de y representam expressão normalizada em relação à população analisada.

#### 4.5.2 TcRBP7 e TcRBP40 apresentam a mesma localização celular

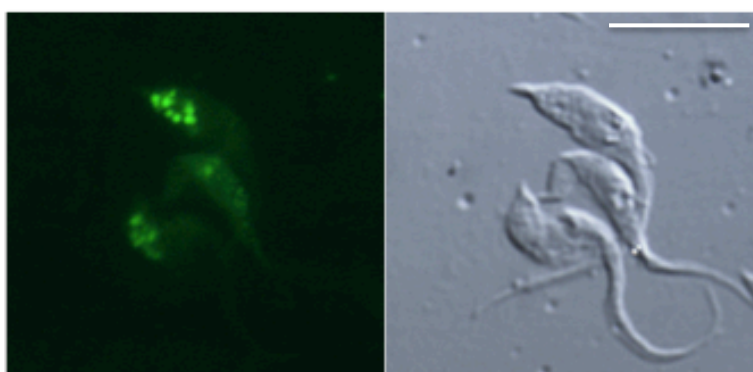
Os genes codificantes das proteínas TcRBP7 e TcRBP40 foram clonados no vetor de fusão carboxi-terminal à proteína GFP, para transfecção em *T. cruzi* e expressão epissomal da fusão. O ensaio de *Western blot* confirmou a expressão das proteínas fusionadas, apresentando a massa molecular prevista para a fusão (Fig. 29).



**Figura 29.** Expressão de TcRBP7 e TcRBP40 fusionadas à GFP. Numeração dos poços respectivamente. Valores expressos em quiloDaltons. A marcação (\*) indica as proteínas fusionadas, observadas em cerca de 45 kDa, resultantes da fusão das proteínas de aproximadamente 15 kDa com a GFP de 30 kDa.

A proteína TcRBP40 foi confirmada estar presente no interior dos reservossomos, conforme demonstrado na seção 4.1 dos resultados.

A localização celular da proteína TcRBP7 foi determinada pela sua fusão à proteína GFP no parasita, cuja eficiência em manter a localização celular foi confirmada para a fusão TcRBP40-GFP na Figura 8. Desta forma, a proteína TcRBP7 também foi vista em *foci* citoplasmáticos específicos na região posterior do parasita, a localização celular dos reservossomos (Fig. 30).



**Figura 30.** Fluorescência de parasitas TcRBP7-GFP. Culturas transfectantes com a fusão C-terminal TcRBP7-GFP aderidas em lâmina para fluorescência direta (painel à esquerda). À direita, parasitas no contraste diferencial de interferência (DIC). Barra: 5  $\mu$ m.

#### 4.5.3 Os genes de *TcRBP7* e *TcRBP40* são parálogos

Frutos de duplicações de um gene de RBP7 ancestral, os genes *TcRBP7* e *TcRBP40* são parálogos que estão assimetricamente divergindo no organismo. Suas proteínas compartilham 72% de identidade, sendo a maior parte desta identidade concentrada dentro do domínio RRM e na região N-terminal.

Cabe agora determinar a atividade destas proteínas, a fim de comparar com as evidências acima apontadas de que as proteínas poderiam vir a ser redundantes, já que estruturalmente e no perfil celular são semelhantes.

A proposta para este trabalho é caracterizar a rede de alvos regulada por estas duas RBPs de *T. cruzi*, por se tratar de um trabalho que já estava em andamento para a proteína *TcRBP40*. Além disso, sendo esta dupla de proteínas mais divergente estruturalmente comparando-se com a dupla de *T. brucei*, o resultado funcional comparativo entre elas já poderia sugerir o resultado para as proteínas de *T. brucei*. Em caso de redundância entre as proteínas do *T. cruzi*, é bem provável encontrar o mesmo fenômeno ocorrendo entre as duplicações em *T. brucei*, que são ainda mais conservadas.

#### 4.5.4 Ribonômica in vivo

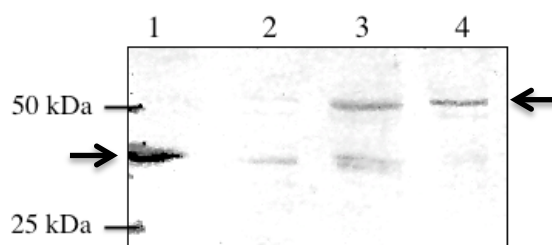
Ribonômica é a identificação em larga escala dos alvos de uma proteína de união ao RNA (TENENBAUM et al., 2000). Diversas abordagens já foram empregadas neste estudo. Para a obtenção dos alvos podem ser utilizadas as estratégias de imunoprecipitação, captura de RNAs *in vitro* com RBP recombinante ou captura por afinidade de proteínas do complexo fusionada a etiquetas. Os alvos são comumente identificados por microarranjo, mas atualmente vêm sendo identificados por sequenciadores de nova geração (SANCHEZ-DIAZ; PENALVA, 2006).

*TcRBP7* e *TcRBP40* são proteínas que conservam mais de 70% de identidade na sua sequência aminoacídica e uma estrutura proteica bem semelhante. Por esta razão, produzir anticorpos policlonais de reconhecimento específico para cada uma delas é inviável. Assim, optou-se neste estudo por trabalhar com a proteína homóloga fusionada a etiqueta de purificação.

A expressão episomal do vetor não garante a correta taxa de expressão e localização das proteínas. Por isso, as imagens de fluorescência anteriormente apresentadas confirmam a localização das proteínas, e sua função original possivelmente é mantida nos transfectantes (Figs. 8 e 30).

#### 4.5.4.1 Imunoprecipitação

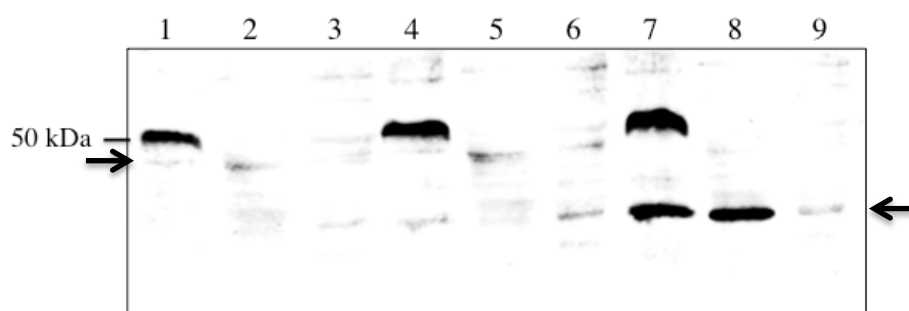
Para capturar os complexos *in vivo*, as culturas transfectantes contendo a fusão TcRBP7-GFP e TcRBP40-GFP foram submetidas a testes de lise, juntamente com a cultura GFP-controle, a fim de obter um extrato solúvel contendo as proteínas fusionadas e capaz de manter os complexos protéicos para o ensaio de imunoprecipitação (IP). A melhor condição foi a que empregou o tampão IMP1, tampão hipotônico contendo NP-40 a 0,5%, em 2 horas de lise. A Figura 31 mostra a fração solúvel das culturas e a presença da marcação contra GFP para GFP-controle e TcRBP7-GFP.



**Figura 31.** Teste de lise solúvel das culturas transfectantes com a fusão GFP C-terminal. *Western blot* com anti-GFP de culturas com a fusão TcRBP7-GFP (3 e 4) e GFP-controle (1 e 2), submetidas a lise com tampão IMP1 e fracionadas. 1 e 3 contêm a fração solúvel, 2 e 4 o precipitado da lise. As setas indicam a posição das proteínas fusionadas: à esquerda, GFP-controle; à direita, TcRBP7-GFP.

O anticorpo anti-GFP foi ligado às esferas magnéticas por imunoafinidade, as quais foram incubadas com o extrato solúvel dos parasitas transfectantes. Após lavagens, as frações ligada e não-ligada foram analisadas por *Western blot*, para verificar a presença das fusões na fração da resina (Fig. 32). A quantidade de proteínas fusionadas a GFP associadas às esferas foi baixa para este teste, conforme visto pela fraca banda indicada pela seta à esquerda. No entanto, a proteína fusionada

é de fato menos expressa do que o controle (seta da direita). Além disso, no teste de IP foi utilizada uma menor quantidade da resina e do anticorpo em relação ao utilizado no experimento final.



**Figura 32.** Imunoprecipitação das fusões a GFP. *Western blot* contra frações da imunoprecipitação das culturas TcRBP7-GFP (4, 5, 6), TcRBP40-GFP (7, 8, 9) e GFP-controle (1, 2, 3) com anti-GFP. Poços 1, 4 e 7 com proteínas extraídas da resina; poços 2, 5 e 8 com a fração não-ligada de cada resina; poços 3, 6 e 9 com as proteínas insolúveis precipitadas de cada lisado, indicando que a maioria das proteínas fusionadas ficou no sobrenadante, que foi submetido à imunoprecipitação. As setas indicam a posição das proteínas contendo GFP imunoprecipitadas, que são as bandas fracas abaixo do anticorpo (o anti-GFP da imunoprecipitação é representado pela banda forte em 50 kDa nas frações da resina 1, 4 e 7). O controle GFP é mais fortemente expresso (7 e 8) do que as proteínas fusionadas.

O ensaio de imunoprecipitação pela etiqueta fusionada foi realizado para culturas contendo as fusões TcRBP7-GFP, TcRBP40-GFP e GFP-controle, cultura transfectada com o vetor sem fusão, em triplicatas biológicas. Os RNAs foram extraídos e sequenciados pela plataforma SOLiD™ 4 System, para identificação dos alvos das proteínas. Um total de 18 amostras foi processado para a ribonômica *in vivo*: frações de RNAs ligado e não-ligado de 2 proteínas mais o controle, todas em triplicatas.

#### 4.5.4.3 Análise do sequenciamento e qualidade

A preparação das amostras para este sequenciamento tem como etapa inicial a fragmentação do RNA. Aos fragmentos são acoplados adaptadores em ambas as extremidades, necessários para a posterior ligação a esferas e amplificação na PCR de emulsão. Os fragmentos foram pré-selecionados por tamanho ainda na etapa de



preparação, sendo escolhidos aqueles entre 150 e 250 nucleotídeos, constituídos pela sequência do RNA fragmentado, com aproximadamente 50 bases, mais os adaptadores.

Após o processo do sequenciamento, os milhões de fragmentos sequenciados, as chamadas leituras, são processadas para a remoção virtual das sequências dos adaptadores. Este processamento é denominado *trimming*. Os fragmentos sequenciados que apresentaram problemas na identificação dos adaptadores foram triados e removidos da contagem de leituras final. A Tabela 3 mostra o número de leituras obtidas para cada amostra, antes e após o *trimming*, permanecendo cerca de 98% das leituras para a etapa seguinte.

As leituras restantes foram submetidas ao mapeamento frente ao genoma de *T. cruzi*, cepa CL Brener. Das leituras restantes, apenas cerca de 23% foram atribuídas aos genes do organismo, incluindo os genes de RNAs não-codificantes, como rRNAs, tRNAs, etc. (Tabela 3).

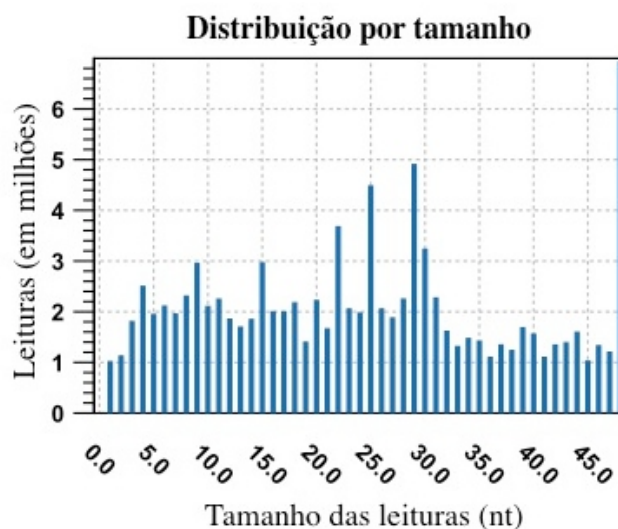
**Tabela 3.** Número de leituras por amostra, pré- e pós-*trimming* e representação do total pela porcentagem. O número à frente do nome das RBPs indica a réplica biológica.

<b>Amostra</b>	<b>Pré- trimming</b>	<b>%pré/ total</b>	<b>Pós- trimming</b>	<b>%pré/ pós</b>	<b>Mapeadas</b>	<b>%map/ pós</b>
RBP7-1	32.014.867	6,6	30.581.629	95,5	6.340.866	21
NL RBP7-1	29.531.035	6,1	29.195.280	98,9	7.511.827	26
RBP7-2	24.668.589	5,1	24.236.765	98,2	5.545.682	23
NL RBP7-2	22.242.678	4,6	22.003.507	98,9	5.281.342	24
RBP7-3	27.837.697	5,8	27.475.667	98,7	4.108.149	15
NL RBP7-3	34.975.709	7,2	34.548.319	98,8	7.656.509	22
RBP40-1	21.510.860	4,4	21.243.650	98,8	3.297.380	16
NL RBP40-1	28.385.190	5,9	28.085.335	98,9	7.059.175	25
RBP40-2	27.617.286	5,7	27.209.405	98,5	5.839.611	21
NL RBP40-2	22.578.653	4,7	22.329.257	98,9	5.696.366	26
RBP40-3	21.395.123	4,4	21.147.345	98,8	1.370.214	6
NL RBP40-3	36.153.564	7,5	35.748.185	98,9	8.167.179	23
Ctrl-1	20.840.033	4,3	20.607.344	98,9	2.163.745	10
NL Ctrl-1	26.785.418	5,5	26.480.887	98,9	6.653.561	25
Ctrl-2	21.958.944	4,5	21.684.500	98,8	2.229.295	10
NL Ctrl-2	28.278.298	5,8	27.957.640	98,9	6.883.088	25
Ctrl-3	22.318.095	4,6	22.035.975	98,7	4.301.632	20
NL Ctrl-3	34.834.446	7,2	34.456.832	98,9	7.330.862	21
RBP7-1b	30.310.692	6,3	29.976.374	98,9	6.152.670	21
RBP40-1b	23.492.731	4,9	23.188.046	98,7	4.095.665	18
<b>TOTAL</b>	<b>483.926.485</b>	<b>100</b>	<b>477.027.522</b>	<b>98,57</b>	<b>107.684.818</b>	<b>23</b>

NL: fração de RNAs não-ligados.

b: repetição de duas amostras para totalizar 20, conferindo equilíbrio à divisão de *barcodes* para o sequenciamento.

O baixo índice de mapeamento das leituras no genoma do organismo pode ser devido à qualidade das sequências. A Figura 33 mostra que cerca de 40% das sequências apresentaram tamanho inferior a 20 nucleotídeos após o *trimming*. Porém, outra possível causa poderia ser a diferença entre as cepas de *T. cruzi* utilizadas neste trabalho: Dm28c para os experimentos e CL Brener para o mapeamento. É possível que o número de polimorfismos presentes nas leituras os tenha excluído do mapeamento. A razão de utilizar o genoma de CL Brener como referência para o mapeamento deve-se ao fato de ser este o genoma melhor anotado do parasita disponibilizado em bases de dados.



**Figura 33.** Distribuição das leituras por tamanho. Gráfico gerado por análise de qualidade pelo software CLC Genomics para uma das amostras. nt: nucleotídeos.

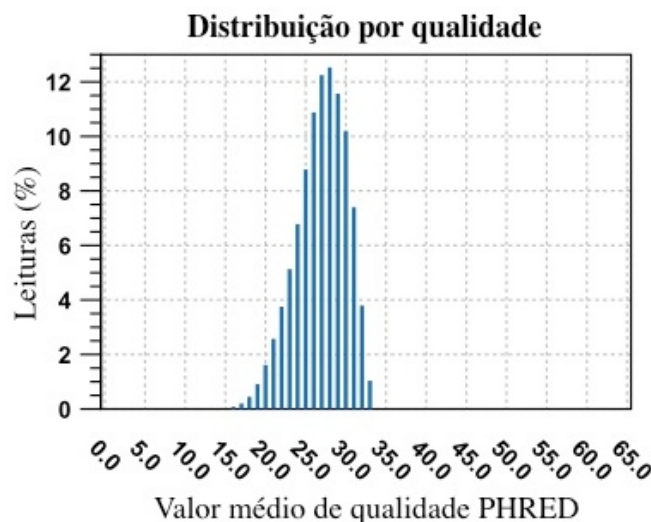
A média do tamanho dos fragmentos nas amostras foi de aproximadamente 22 nucleotídeos, quando a expectativa seria de 45. Além disso, cerca de 5 milhões de leituras apresentaram valor inferior a 5 nucleotídeos, que podem ser atribuídos a qualquer sequência no mapeamento.

O programa realizou uma análise de qualidade dos fragmentos sequenciados. Esta análise avalia a confiabilidade da identificação de cada base sequenciada, dada pelo valor de qualidade da base. A métrica mais utilizada é o Valor de Qualidade PHRED, calculado segundo a fórmula  $VQP = -10 \cdot \log_{10}(p)$ , onde  $p$  é a probabilidade de a base ter sido erroneamente identificada (EWING et al., 1998). O valor resultante pode ser assim interpretado segundo a Tabela 4.

**Tabela 4.** Relação entre o Valor de Qualidade PHRED e a probabilidade de erro na determinação de uma base.

Valor de PHRED	Probabilidade de erro	Precisão
10	1 em 10	90%
20	1 em 100	99%
30	1 em 1000	99,9%
40	1 em 10000	99,99%
50	1 em 100000	99,999%

Neste sequenciamento, a maioria das leituras ficou com valor de qualidade entre 25 e 30, indicando 99,9% de precisão (Fig. 34). Este valor é inferior ao observado para outros sequenciamentos realizados no mesmo equipamento. Isso significa que as bases foram atribuídas com confiança inferior ao esperado, podendo haver erros na determinação das sequências dos fragmentos.



**Figura 34.** Distribuição do Valor de Qualidade PHRED entre as leituras de uma das amostras.

Uma vez que o mapeamento das leituras no genoma tenha obedecido ao critério de no máximo 1 diferença de nucleotídeo no pareamento, a baixa qualidade das leituras também explica a baixa proporção de leituras mapeadas. Apesar disso, a quantidade de leituras mapeadas indicada na Tabela 3 atende à demanda de uma ribonômica, a qual não requer cobertura do genoma todo, uma vez que o número de alvos regulados pelas RBPs é em média de 100 genes transcritos (HOGAN et al., 2008) e portanto 1 milhão de leituras seria mais do que o necessário para a cobertura dos alvos de cada RBP. Posto isso, a análise de determinação dos alvos das RBPs em estudo permanece válida.

#### 4.5.4.4 Seleção dos alvos de TcRBP7 e TcRBP40

As leituras acima mencionadas são provenientes da fragmentação dos alvos dos complexos RNP formados pelas proteínas TcRBP7 e TcRBP40. O mapeamento

permitiu determinar a quais genes pertencem os fragmentos. A lógica é de que os transcritos mais abundantes nos complexos apresentem maior número de fragmentos, considerando-se o tamanho do transcrito (gene) e o total de fragmentos obtidos na amostra. Este valor é o RPKM (*reads per kilobase per million*), cujo índice normalizado foi utilizado para determinar o enriquecimento das amostras de IP em relação aos seus controles.

Para a determinação dos alvos de cada RBP dentre os genes identificados foram adotados 5 filtros com critérios arbitrários de seleção. A aplicação de cada filtro causou um corte no total de genes das amostras, representados na tabela 5.

O primeiro filtro de seleção foi a presença de leituras (valor RPKM > 0) nos três experimentos (triplicatas) de cada proteína. O número de réplicas biológicas foi novamente considerado no terceiro filtro.

O segundo filtro determinou um mínimo de 5 leituras na média das triplicatas do RNA ligado em cada RBP.

O terceiro critério foi em relação ao enriquecimento dos transcritos na fração ligada. Para isso, cada réplica da IP foi comparada à sua fração não-ligada correspondente. O enriquecimento de cada gene foi determinado pela razão do seu RPKM nestas duas frações, o que é denominado *fold change* (FC) do gene. Assim, os genes considerados enriquecidos na IP de cada RBP foram aqueles que apresentaram média das razões igual ou superior a 2 ( $\overline{FC} \geq 2$ ), o que significa um enriquecimento mínimo de 2 vezes no precipitado em relação à população de transcritos não ligados.

O quarto filtro determinou que pelo menos duas das réplicas deveriam ter o valor de FC igual ou superior a 2, desde que a terceira réplica fosse também enriquecida no precipitado, ainda que com valor inferior a 2. Apesar de não empregar valores estatísticos, este critério de seleção trouxe apenas os genes cujos valores seriam considerados significativos.

O quinto filtro, por fim, considerou os valores apresentados na IP-controle, realizada com a proteína GFP sem fusão a outra proteína de *T. cruzi*. Este controle determina a exclusão dos transcritos (ou complexos contendo RNAs) que porventura se ligaram inespecificamente aos próprios instrumentos do experimento, como as esferas magnéticas e a proteína GFP. O valor do enriquecimento nos experimentos com as RBPs foi determinado pela diferença entre as médias de FC dos alvos das

RBP7 e do controle ( $\Delta\overline{FC} = [\overline{FC} \text{ da RBP}] - [\overline{FC} \text{ do controle}]$ ), considerando os alvos 2 vezes ou mais enriquecidos nas RBP7. Neste momento, qualquer valor dos genes do controle foi considerado no cálculo da diferença, independente se o gene passou ou não pelos filtros anteriores na amostra controle. Os filtros anteriores foram apenas empregados à amostra controle para fins de cálculo comparativo com os números de alvos restantes para cada RBP (conforme a Tabela 5).

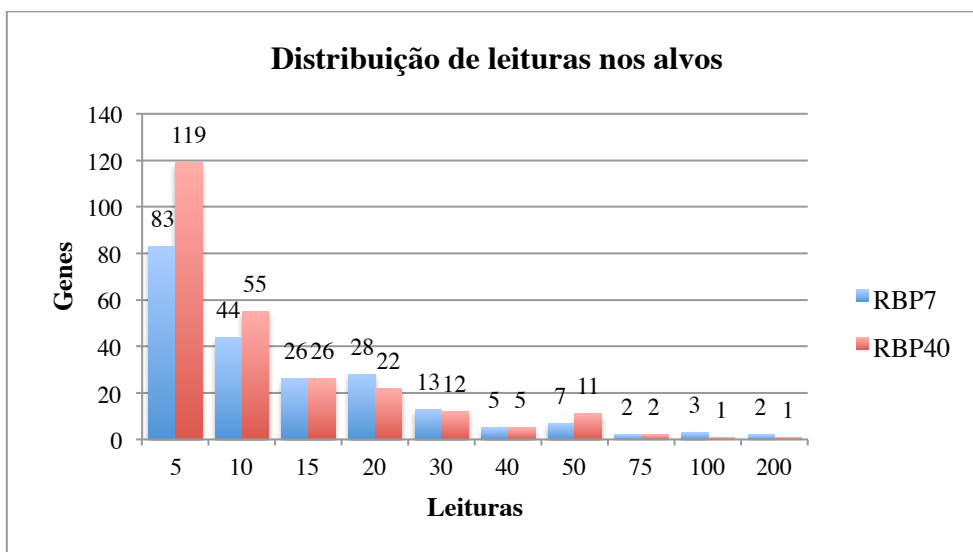
**Tabela 5.** Quantidade de alvos restantes após a aplicação de cada filtro de seleção, considerando um valor inicial de 25.211 genes. A última linha representa o número final de alvos para cada RBP.

<b>Filtros</b>	<b>RBP7</b>	<b>RBP40</b>	<b>Controle</b>
triplicata e RPKM > 0	11121	9968	9263
$\geq 5$ leituras	8669	6994	6025
$\overline{FC} \geq 2$	2797	2326	2075
$FC \geq 2$ em 2 ou 3 réplicas	901	1002	925
$\Delta\overline{FC} \geq 2$	213	254	-

FC: *fold change*.

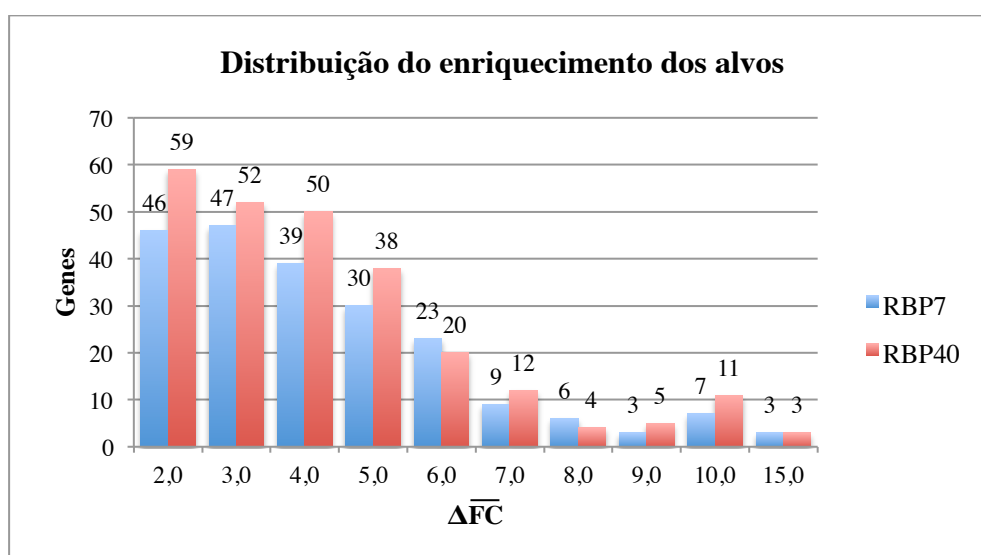
A lista completa dos alvos das proteínas TcRBP7 e TcRBP40 está disponível no Anexo, incluindo os valores de cada critério de seleção e a anotação dos genes, quando existente.

Foram calculadas as distribuições dos valores de seleção de cada gene dentre os alvos (genes que passaram pelos 5 filtros), buscando verificar a representatividade dos mesmos. Quanto ao número de leituras em cada gene, a maioria dos alvos das duas proteínas apresentou a média entre 5 e 15 leituras (Fig. 35). A soma das médias das leituras foi de 6.176 para RBP7 e 5.273 para RBP40, uma diferença de 14,6% a mais em RBP7.



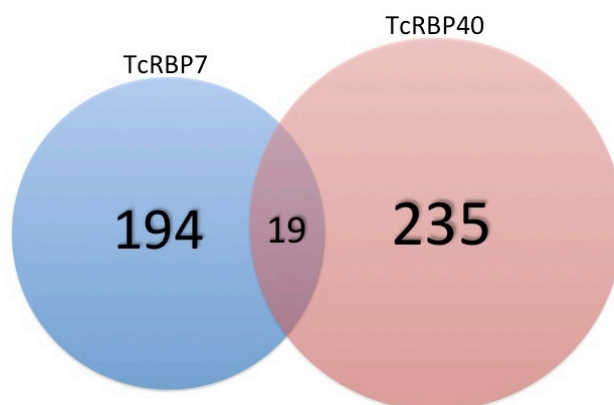
**Figura 35.** Número de leituras (fragmentos sequenciados mapeados) para cada alvo das RBPs. O eixo x indica o número mínimo de leituras dentro da classe, até o limite do valor seguinte. Os números acima de cada barra indicam a contagem de alvos (valor do eixo y).

A seguir foram analisados os valores do enriquecimento dos alvos das RBPs em relação aos valores apresentados pelo controle, referentes ao último critério de seleção (Fig. 36). Mais de 50% dos genes têm valor de enriquecimento até 5.



**Figura 36.** Distribuição do valor de enriquecimento dos genes em relação ao controle. O eixo x indica o valor mínimo do enriquecimento ( $\Delta\overline{FC}$ ) dentro da classe até o limite do valor seguinte.

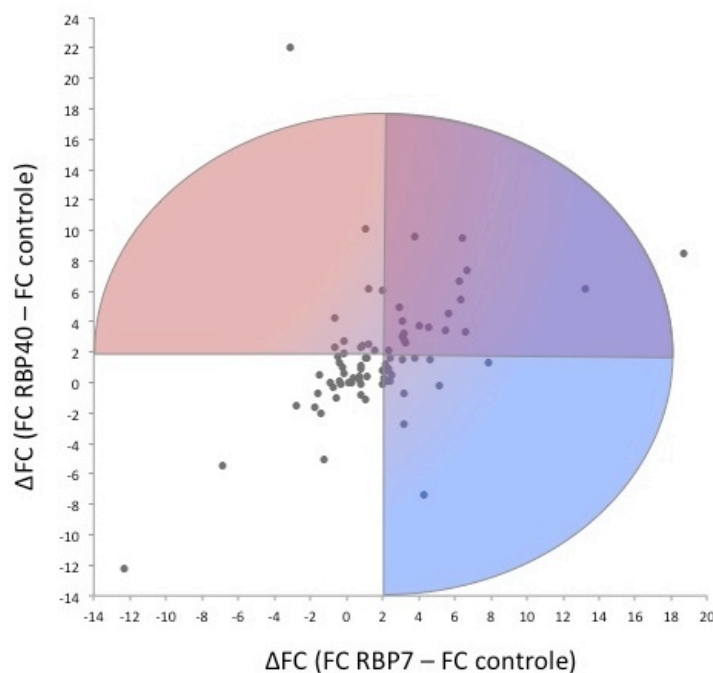
Foram constatados 19 alvos em comum entre as duas proteínas, utilizando todos os filtros de seleção para ambas simultaneamente (Fig. 37), correspondendo a aproximadamente 7% do total de alvos de cada RBP.



**Figura 37.** Gráfico de Venn contabilizando os alvos de TcRBP7 e TcRBP40. A área em azul contém o número de alvos de TcRBP7 e em vermelho de TcRBP40. A área de interseção indica o número de alvos em comum entre as duas proteínas.

Foi realizada uma análise dos valores de todos os genes que se encaixaram nos quatro primeiros critérios para ambas as RBPs, ou seja, todos os valores possíveis para  $\Delta\overline{FC}$  das duas RBPs. A Figura 38 mostra o valor de cada um destes genes em relação às duas amostras, TcRBP7 e TcRBP40. Há vários genes próximos às áreas delimitadas, seja na porção interna ou externa das áreas. Isso mostra que há alguns genes no limite da linha de corte final, e portanto poderiam ser excluídos ou incluídos na lista final de alvos, restringindo-se ou afrouxando-se o critério de seleção.





**Figura 38.** Dispersão dos genes em razão de TcRBP7 e TcRBP40. Os valores  $\Delta FC$  são a diferença do enriquecimento em relação ao controle. As meia-luas em azul e vermelho indicam os alvos positivos de cada proteína, respectivamente, e a região em roxo é a área de interseção, correspondente aos alvos em comum. Os pontos externos às áreas destacadas, mas acima das linhas de corte, são valores extremos dos alvos de uma ou outra proteína.

Os gráficos de distribuição acima apresentados (Figs. 36 e 37) também demonstram esta mesma conclusão, de que restringir os filtros de seleção implica em excluir uma determinada quantidade de alvos. Na tabela do Anexo é possível identificar quais são estes genes próximos aos limites de seleção.

Deste modo, as análises das funções dos genes e especulações funcionais das redes de cada proteína seguirão este padrão. Uma posterior análise poderá extrapolar para um padrão mais “relaxado” de seleção e verificar se a inferência funcional das redes é alterada.

#### 4.5.5 Ribonômica *in vitro*

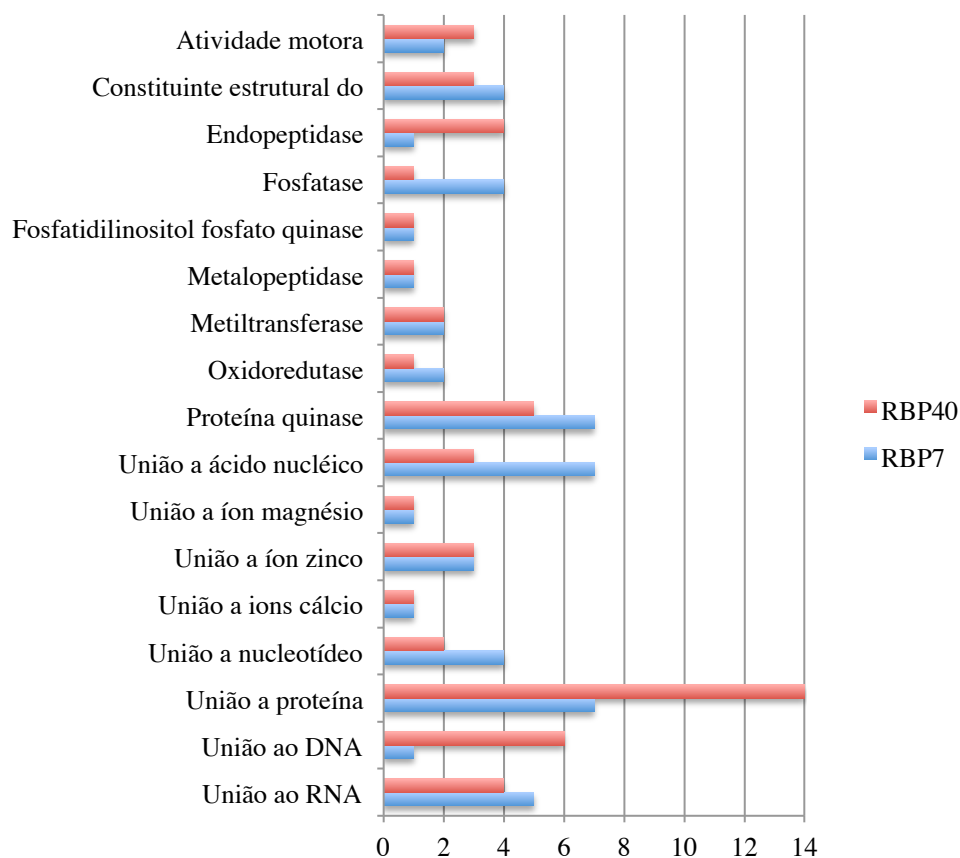
Os mRNAs aos quais as proteínas TcRBP7 e TcRBP40 se ligam foram capturados por ribonômica *in vivo* e identificados por sequenciamento em massa.

A primeira validação experimental utilizou uma metodologia ribonômica *in vitro*, onde os mRNAs alvos das proteínas foram capturados por *pull-down* de RNAs, utilizando a proteína recombinante purificada como isca. Os RNAs associados foram também identificados por sequenciamento em massa na plataforma SOLiD™ 4 System, e estão atualmente em análise.

#### 4.5.6 Análise funcional dos alvos de TcRBP7 e TcRBP40

Do total de alvos designados para cada proteína, apenas cerca de 33% têm anotação funcional, baseado na homologia com genes anotados em outros organismos. Este é um problema comumente enfrentado pelos grupos de estudos em *T. cruzi* envolvendo análises genômicas, visto que a maioria dos genes não têm função conhecida (EL-SAYED et al., 2005).

Ainda assim, os genes alvos de TcRBP7 e TcRBP40 com anotação disponível foram analisados de acordo com sua função molecular, utilizando a terminologia de ontologia gênica (HUNTLEY et al., 2014), disponível no banco de dados TriTrypDB 8.0. Vários termos foram comuns para os alvos das duas proteínas (Fig. 39).



**Figura 39.** Função molecular em comum dentre os alvos de TcRBP7 e TcRBP40. Os valores em x indicam o número de alvos contendo os termo apresentado em y.

Alguns termos apenas apareceram nos alvos de uma das proteínas. TcRBP40 teve o maior número de termos únicos, comparando com TcRBP7 (Tabs. 6 e 7).

**Tabela 5.** Função molecular exclusiva dos alvos de TcRBP7.

Função Molecular	RBP7
Ácido graxo elongase	1
Atividade com NADP+	3
ATPase	1
Descarboxilase	1
Formilmetionina deformylase	1
GTPase	3
Ligase de cadeia longa de CoA ácido	1
Polinucleotídeo 3'-fosfatase	1
Tubulina-tirosina ligase	1
Ubiquitina-ligase	1
União a íon ferro	2
União a ribossomo	1

**Tabela 6.** Função molecular exclusiva dos alvos de TcRBP40.

<b>Função Molecular</b>	<b>RBP40</b>
Aspartato carbamoiltransferase	1
Ativador de GTPase	1
Atividade catalítica	5
Carreador de elétrons	1
Constituinte estrutural do citoesqueleto	2
Diacilglicerol quinase	1
Fator de iniciação traducional	1
Fosfodiesterase de nucleotídeo 3',5'-cíclico	1
Glicilpeptídeo N-tetradecanoiltransferase	1
Hidrolase de ester/diester fosfórico	2
Liase	1
Ligase de aminoácidos ácidos	1
Molécula estrutural	2
NADH desidrogenase (ubiquinona)	2
Peptídeo transferase	1
poli(ADP-ribose) glicoidrolase	1
Proteína de ancoramento de NSF solúvel	1
Queuino tRNA-ribosiltransferase	1
Receptor de SNAP	1
RNA polimerase guiada por DNA	1
Transferase	2
Transportador de íons de cálcio pela membrana	1
Transportador de membrana de íon metálico	1
Transportador de membrana de piruvato	1
Triglicerídeo lipase	1
União a actina	1
União a aminoácido	1
União a DNA danificado	1
União a fosfolipídeo	2
União a GTP	1
União a heme	1
União a proteína de choque térmico	1

Em ambos os casos, há um grande número de alvos codificando proteínas envolvidas na ligação de íons e outras moléculas, atividades de transporte entre membranas e enzimas do metabolismo de diversas vias.

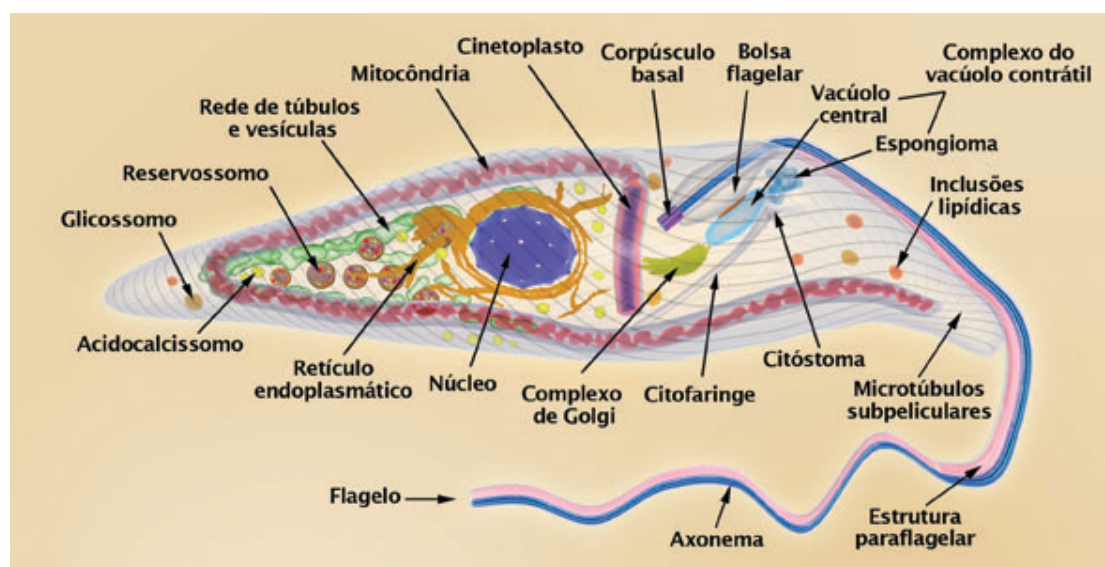
Futuras análises permitirão enquadrar as diferentes proteínas em vias comuns, no objetivo de determinar as redes funcionais reguladas pelas RBPs.

## 5 DISCUSSÃO

### 5.1 Localização da TcRBP40 e os Reservossomos

A proteína TcRBP40 foi alvo de estudos anteriores (GUERRA-SLOMPO et al., 2012). Sua investigação se iniciou por ser esta uma proteína contendo o domínio RRM de interação com RNAs e que se mostrou exclusiva em *T. cruzi*. Como proteína regulatória da expressão gênica, poderia atuar em vias que são exclusivas do organismo, sendo uma das proteínas que determinariam as diferenças apresentadas entre as espécies mais semelhantes.

Experimentos preliminares haviam mostrado que sua localização estaria nos reservossomos das formas epimastigotas do parasita, o que chamou atenção por ser esta uma organela de armazenamento nutricional, abrigando proteases e fatores envolvidos com o metabolismo do parasita, como destino final da via endocítica (Fig. 40).



**Figura 40.** Epimastigota de *T. cruzi*. Ilustração esquemática mostrando as estruturas celulares do parasita. Fonte: Atlas Didático – O ciclo de vida do *Trypanosoma cruzi*, LabSpace.

No presente trabalho foi confirmada esta localização pelo uso de outras metodologias, como a fusão a etiquetas de identificação, a co-localização com partículas endocitadas de transferrina e a obtenção de uma fração enriquecida de

reservossomos. Desta forma, as especulações a seguir se dão tendo por certo que elementos regulatórios, reconhecedores de mRNAs, fazem parte do proteoma desta organela.

Não obstante, foi constatada a localização celular do mRNA codificante da proteína TcRBP40 dentro do mesmo compartimento, pela metodologia de hibridização *in situ*. Este mesmo fenômeno da localização de mensageiros próximo ao local das suas proteínas foi reportado também em outros exemplos, como a B-tubulina na região perinuclear, PFR2 (*Paraflagellar Rod 2*) próximo da região da bolsa flagelar, e mesmo a cruzipaína, peptidase abundante dos reservossomos (ALVES et al., 2013). Conclui-se que o parasita *T. cruzi* apresenta localização diferencial de mRNAs, em concordância com a localização celular das proteínas por eles codificadas.

De cada um destes genes foi verificado, pelos ensaios de localização de gene repórter, que a sua região regulatória à 3' contém o motivo para localização diferencial do mensageiro, e também no mRNA da TcRBP40. Estas sequências regulatórias em *cis* nos mRNAs são denominadas *zipcodes*, pois contêm o sinal de endereçamento para transporte pelo citoplasma. São sequências que podem ser conservadas entre várias espécies, como é o caso do mRNA da  $\beta$ -actina, que tem o motivo hexanucleotídico ACACCC e destina estes mensageiros para regiões onde há lamelipódios em alguns tipos celulares, conservado em células de galinha e humanas (KISLAUSKIS; ZHU; SINGER, 1994). Estes elementos estão preferencialmente localizados nas regiões 3'UTR dos mensageiros e contêm entre 5-6 a até centenas de nucleotídeos (MARTIN; EPHRUSSI, 2009).

O reconhecimento dos *zipcodes* é realizado por RBPs. No exemplo acima citado, o reconhecimento e destinação são feitos pela proteína ZBP1, que contém dois domínios RRM e 4 K-homology (FARINA et al., 2003; ROSS et al., 1997). Foi visto que são estes domínios RRM os responsáveis pela destinação dos complexos RNP para o local correto na célula, provavelmente os que fazem a conexão com as proteínas motoras dineína e cinesina.

Outro exemplo clássico de endereçamento de mensageiros é nos oócitos de *Drosophila*, onde a presença diferencial de mRNAs que virão a compor as diferentes estruturas corporais do inseto é controlada por RBPs específicas. Foi observado que

as proteínas Vsp22, Vsp25 e Vsp36 (*Vesicular Sorting Protein*), componentes do complexo ESCRT II (*Endosomal Sorting Complex Required for Transport*), fazem conjuntamente o reconhecimento do mRNA *bicoid* e promovem sua destinação até o polo anterior do oócito (IRION; ST JOHNSTON, 2007). Este e outros dois complexos proteicos conservados evolutivamente, ESCRT I e III, fazem a triagem de proteínas de membrana presentes nos endossomos, destinando-as para outros compartimentos celulares, como para degradação, retorno à membrana plasmática, secreção, etc. (HURLEY; EMR, 2006). Outras influências na expressão gênica foram reportadas para estas proteínas, ainda que os mecanismos não tenham sido desvendados (SLAGSVOLD et al., 2006). As proteínas deste complexo foram assim reportadas como sendo “*moonlighting*”, pois exercem funções diferentes da sua conhecida principal. Esta dupla função de algumas proteínas pode ter se originado a partir de uma pressão evolutiva, que causou a seleção de proteínas e complexos multitarefas, minimizando assim o número de genes necessários para compor os organismos (RUSTEN; STENMARK, 2007).

Uma análise preliminar utilizando um programa de predição de motivos conservados em sequências mostrou que há várias porções nas regiões intergênicas à jusante dos genes TcRBP40 e cruzipaína com potencial de ser um elemento de destinação dos mensageiros para os reservossomos, dada sua conservação e características que os enquadrariam como *zipcodes*. Futuras análises utilizando estas sequências em genes repórteres e ensaios de interação poderão confirmar a correta sequência. Além destes ensaios, um trabalho em colaboração com a Dra. Beatriz Garat e a MSc. Lucía Pastro, do Laboratorio de Interacciones Moleculares, da UdelaR, Uruguay, está sendo realizado para identificação dos RNAs presentes nos reservossomos. Os RNAs sequenciados obtidos de amostras enriquecidas de reservossomos estão em etapa de análise. Após sua identificação, será possível buscar por elementos em comum dentre os RNAs confirmadamente presentes nos reservossomos.

Constatando a presença de mRNAs nos reservossomos, surge o questionamento se haveria tradução na organela. No trabalho de Alves e colaboradores, a fração enriquecida de reservossomos foi submetida a lise da organela e análise dos RNAs ali contidos. Os rRNAs não são vistos na fração lisada, que detém apenas o conteúdo

interno dos reservossomos (ALVES et al., 2013). Outro trabalho realizou a ultra-estrutura do *T. cruzi* e detectou uma distribuição desigual de ribossomos no citoplasma, mostrando uma grande concentração deles no entorno dos reservossomos (GIRARD-DIAS et al., 2012). Isto mostra que a região posterior do parasita é traducionalmente ativa. No estudo não foram observados ribossomos no interior da organela.

Outro achado de elementos regulatórios nos reservossomos foi reportado pelo grupo de Garcia-Silva, em 2010, onde foram detectados pequenos RNAs derivados de tRNAs clivados (stRNAs) co-localizando com a cruzipaina. Não se sabe a função destes pequenos RNAs, mas especula-se a participação da proteína TcPIWI-tryp, proteína homóloga à Argonauta de eucariotos superiores, em vias alternativas de regulação de mRNAs (GARCIA-SILVA et al., 2010, 2014c).

Alguns eucariotos apresentam comunicação celular por meio de vesículas exocitadas, as quais abrigam proteínas e RNAs. Em *T. cruzi* foi reportado que vesículas extracelulares secretadas pelo parasita desencadeiam a diferenciação para as formas infectivas e ainda aumentam a susceptibilidade das células de mamíferos à invasão pelo parasita (GARCIA-SILVA et al., 2014a, 2014b). Estas vesículas contêm stRNAs e ainda a proteína TcPIWI-tryp. Estas moléculas foram vistas no interior dos reservossomos, e as vesículas exocitadas brotam da região posterior do parasita. Desta forma, é possível que os reservossomos participem desta via de endocitose/exocitose de vesículas no parasita.

Isto põe em cheque a real ou completa função dos reservossomos para o parasita. Estes achados podem sugerir uma função além da estocagem nutricional de moléculas provenientes da via endocítica e sua reciclagem, sendo também um local de triagem de moléculas destinadas a vesículas exocitadas para a comunicação intercelular. Assim, os reservossomos não seriam mais vistos como um “depósito”, mas um “porto”, gerenciado por proteínas regulatórias das quais TcRBP7 e TcRBP40 poderiam fazer parte.



## 5.2 Filogenia da família RBP7 em tripanosomas

*T. cruzi* e *T. brucei* são espécies distantes geograficamente. Enquanto *T. cruzi* ocupa as regiões tropicais das Américas, tendo já se espalhado pelas regiões no entorno, o *T. brucei* se concentra no continente africano. Seus ciclos de vida envolvem também hospedeiros artrópodes e mamíferos, mas os mecanismos de entrada e infecção são distintos. Enquanto o *T. cruzi* invade células e ali se replica, o *T. brucei* apenas ocupa a corrente sanguínea do hospedeiro mamífero (KRAMER, 2012).

O genoma do *T. cruzi* é considerado altamente repetitivo (pelo menos 50%), regiões estas consistindo principalmente famílias multigênicas que codificam proteínas de superfície, retrotransposons e regiões sub-teloméricas. Estas expansões gênicas são consideradas recentes, devido à sua alta conservação, ou podem ser mantidas conservadas por mecanismos específicos de contração nestas expansões. Um exemplo é a família de miosinas em *T. cruzi*, expandida apenas dentro da espécie, mas com uma considerável diferença dentre estes genes, já em diferentes loci (EL-SAYED et al., 2005).

Já foi anteriormente ressaltada a importância das RBPs para a regulação pós-transcricional da expressão gênica nestes organismos, que dependem majoritariamente deste nível de regulação para controlar a expressão diferencial dos seus genes. As RBPs concentram grande importância nos processos de diferenciação dos estágios de vida dos parasitas, conforme já demonstrado que ¼ dos genes aumentados no início da transição são codificantes de RBPs (KRAMER, 2012). Estas, por sua vez, respondem a sinais que determinam seus níveis e modo de ação. A proteína TbRBP7B, cuja análise filogenética foi mostrada nesta tese, foi recentemente reportada como sendo parte do sistema de *quorum sensing* em *T. brucei* (MONY et al., 2014). Ela mostrou atuar no ciclo celular do parasita, regulando a parada em G1/G0 típica das formas *stumpy*, o que é dependente da densidade populacional do parasita na corrente sanguínea do hospedeiro vertebrado. Poucos transcritos foram afetados diante da perturbação na expressão gênica da TbRBP7B, dentre eles foram vistos reguladores do RNA, transcritos de prociclinas e de histonas, consistente com os efeitos no ciclo celular e na diferenciação observados. Isso revela a importância

das RBPs em estudo para a biologia destes parasitas, e em especial da família de RBP7 em tripanossomatídeos.

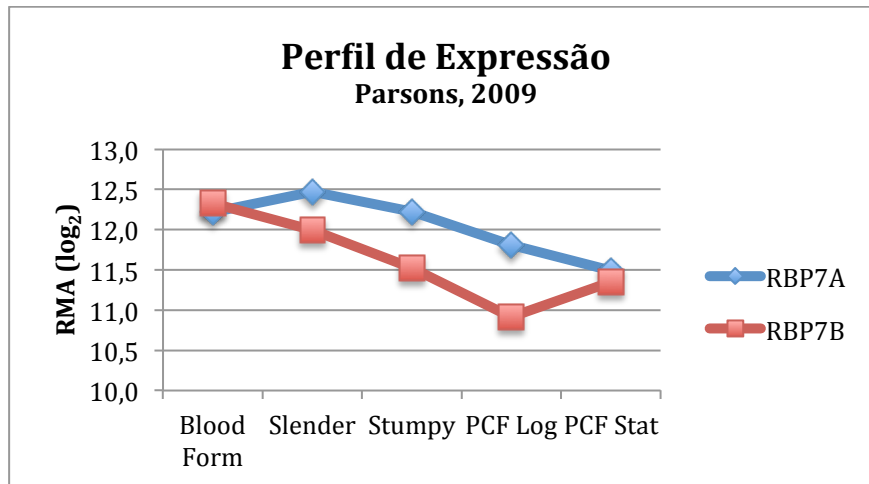
O presente trabalho mostrou a presença de duas proteínas de união ao RNA em *T. cruzi* e *T. brucei* que sofreram eventos de duplicação gênica após a especiação destes organismos. A análise filogenética mostra que os genes duplicados estão acumulando mutações assimetricamente, sendo que aparentemente um deles mantém maior conservação de sua sequência original em cada organismo, a saber, os genes de TcRBP7 em *T. cruzi* e de TbRBP7A em *T. brucei*.

O *T. cruzi* conserva várias cópias idênticas do gene de TcRBP7. Não foi possível determinar por análises comparativas da região das duplicações quantas são as cópias, por ser uma região repetitiva. Também as regiões intergênicas das cópias de TcRBP7 são idênticas, não sendo possível determinar se são reais cópias totalmente conservadas por serem recentes ou de problemas no sequenciamento.

Isso deixa em aberto a possível explicação das duplicações da família, se foram múltiplas e apenas a TcRBP40 divergiu, ou se uma primeira duplicação ocorreu, e mais tarde sucessivas duplicações da TcRBP7 surgiram na região. As várias cópias do gene de TcRBP7 poderiam aumentar a taxa da proteína expressa e conferir vantagem ao parasita. No entanto, estudos de transcritômica e proteômica realizados nos últimos anos ainda não puderam detectar a taxa de expressão da proteína TcRBP7 no organismo (DE GODOY et al., 2012; WEATHERLY; BOEHLKE; TARLETON, 2009). É possível que sua expressão adote o mesmo perfil apresentado pela proteína TcRBP40 – expressa diferencialmente no ciclo de vida –, sustentado pela conservação entre ambas.

A análise filogenética mostra que a duplicação ocorrida em *T. brucei* se manteve mais conservada do que em *T. cruzi*. Há duas explicações evolutivas possíveis: 1) os genes de *T. brucei* estão sob maior pressão seletiva, sem o relaxamento da seleção purificadora, o que permitiria um deles divergir mais em relação ao outro; ou 2) a duplicação seria mais recente e portanto teve menor tempo para o acúmulo de mutações como em *T. cruzi*. Entretanto, análises em larga-escala da expressão gênica em *T. brucei* mostrou que TbRBP7A e TbRBP7B têm diferente perfil de expressão nas diferentes formas do ciclo de vida do parasita. Segundo o trabalho do grupo de Parsons, TbRBP7B teria menor nível de expressão nas formas replicativas, igualando-

se a TbRBP7A nas formas estacionárias ou infectivas (Fig. 41) (JENSEN et al., 2009).



**Figura 41.** Perfil dos transcritos de TbRBP7A e TbRBP7B em diferentes formas do ciclo de vida do *T. brucei*. Comparação dos níveis de RNA de genes codificantes de proteínas e RNAs estruturais em 5 diferentes estágios do ciclo celular: sanguíneo de cultura, sanguíneo alongado *in vivo*, sanguíneo stumpy *in vivo*, cultura procíclica em fase logarítmica, e cultura procíclica em fase estacionária. O eixo x representa a média de três réplicas biológicas independentes dos cinco estágios do ciclo de vida. No eixo y, valores normalizados de RMA (log<sub>2</sub>) ou valor percentual de expressão. Fonte dos dados: (JENSEN et al., 2009)

Se considerada a hipótese de a duplicação em *T. brucei* ser mais recente, estas duplicações em ambos organismos seriam um modelo que comprova a teoria apontada por Koonin, 2009, que sugere que ocorra uma rápida (muito rápida) subfuncionalização seguida de lenta neofuncionalização. Esta seria a primeira evidência capaz de explicar o surgimento de novas proteínas regulatórias. A proposição virá a ser confirmada pelos dados experimentais funcionais das proteínas em estudo.

### 5.3 Divergência funcional de TcRBP7 e TcRBP40

As proteínas TcRBP7 e TcRBP40 tiveram seus mRNAs alvos identificados por ribonômica *in vivo* e sequenciamento de nova geração. Mesmo sendo uma análise

preliminar, foi possível observar que há menos de 10% de sobreposição dos alvos das duas proteínas.

Uma possível explicação para esta divergência está na estrutura molecular destas RBPs. O domínio RRM é muito conservado em estrutura, mas sua sequência primária pode variar entre as proteínas, o que confere as diferenças de afinidade, estabilidade e especificidade pelos alvos. TcRBP7 e TcRBP40 compartilham 72% de identidade na sequência primária. Dos 28% de divergências, parte está dentro do domínio RRM e inclusive nos motivos de reconhecimento RNP-1 e RNP-2, localizados nas fitas  $\beta$ , conforme mostrado na sobreposição dos modelos das estruturas. Além disso, é possível que outras regiões do domínio participem do reconhecimento, como alças e hélices (CLÉRY; BLATTER; ALLAIN, 2008), regiões que também apresentam diferenças de resíduos nestas RBPs.

Alguns estudos mostram que trocas de aminoácidos em posições críticas de RBPs alteram a especificidade do motivo em *cis* que reconhecem, permitindo desenvolver RBPs com determinadas funcionalidades (CHEN; VARANI, 2013). A natureza já faz isso. TcRBP7 e TcRBP40 podem ser o primeiro caso conhecido mostrando que mesmo estando muito próximas filogeneticamente, podem ter um papel bem diferente para a biologia do organismo, regulando redes de alvos distintas.

Outras proteínas envolvidas na regulação pós-transcricional foram identificadas juntamente com seu parálogo em tripanossomatídeos. A proteína DED1 em *Leishmania* duplicou após a ramificação dos organismos. Compartilhando 67% de identidade, mostraram redundância funcional quando complementadas em outro organismo (ZINOVIEV et al., 2012). O complexo de união ao RNA mitocondrial 1 (MRB1) de *T. brucei* teve duas subunidades caracterizadas, as quais constituem parálogos provenientes de uma duplicação exclusiva do organismo, conservando 77% de similaridade. Análises funcionais sugeriram que ambas interagem com as mesmas famílias de RNAs mitocondriais, mas afetam os RNAs pan-editados em grau diferente (KAFKOVÁ et al., 2012).

A fusão a GFP de TcRBP7 e TcRBP40 mostrou mesma localização celular: o interior dos reservossomos do parasita. Isto sugere que o elemento de destinação das proteínas para os reservossomos tenha se mantido conservado. Informação acerca do perfil de expressão em epimastigotas e metacíclicos mostrou um padrão semelhante

entre ambos transcritos, apesar de isso não significar igualdade de taxa de tradução. TcRBP40 foi anteriormente reportada como sendo ausente em metacíclicos (GUERRA-SLOMPO et al., 2012), apesar de seu mRNA estar no parasita e associado aos ribossomos (Fig. 28). Com a ausência de um anticorpo específico para TcRBP7, não foi possível determinar experimentalmente seu perfil de expressão. Futuras análises irão determinar o papel funcional destas proteínas para os seus alvos e para a biologia do parasita.

## CONCLUSÃO

A proteína TcRBP40, anteriormente avaliada como exclusiva do organismo, é na realidade um parálogo da proteína TcRBP7, cujos ortólogos são as também duplicadas TbRBP7A e TbRBP7B de *T. brucei*.

Os eventos de duplicação ocorreram independentemente nos dois organismos, após sua especiação. Os parálogos estão assimetricamente acumulando mutações, sendo TcRBP40 a mais distante do grupo.

TcRBP40 está presente nos reservossomos das formas epimastigotas do parasita, co-localizando com a cruzipaina. Sendo uma proteína atuante na regulação pós-transcricional do parasita, a organela pode desempenhar novas funções não anteriormente exploradas, também regulatórias.

Os parálogos TcRBP7 e TcRBP40, apesar de compartilharem 73% de identidade, se associam a grupos diferentes de mRNAs, com menos de 10% de sobreposição. Isto indica divergência funcional, uma vez que regulam redes distintas de transcritos, afetando assim vias diferentes do organismo.

O conjunto de análises funcionais em *T. cruzi*, mais os trabalhos experimentais também em *T. brucei* que vierem a ser realizados, permitirão uma compreensão acerca da evolução de redes regulatórias nos eucariotos, e assim ter mais um entendimento sobre a complexidade adquirida pelos organismos ao longo das gerações.

## PERSPECTIVAS

O papel das proteínas TcRBP7 e TcRBP40 continuará sob investigação, para que as inferências sobre sua divergência funcional seja completa.

A ribonômica *in vivo* será validada experimentalmente por uma ribonômica *in vitro*, cujo resultado será em breve analisado. Adicionalmente, os alvos identificados serão confirmados por RT-PCR da imunoprecipitação.

Com a confirmação dos alvos, suas funções serão mais profundamente investigadas, buscando determinar a quais redes funcionais as RBPs se associam e regulam. O motivo de reconhecimento será determinado ao buscar por padrões nas regiões regulatórias dos alvos, e será confirmado por ensaios funcionais de interação e competição. Cada RBP será testada com o outro motivo, para verificar se são capazes de fazer reconhecimento cruzado. Os motivos serão clonados em vetores de expressão de genes repórter, para determinar o tipo de regulação desempenhado pelas RBPs.

As culturas com as fusões TcRBP7-GFP e TcRBP40-GFP serão submetidas a metaciclogênese *in vitro*, para determinar a expressão e localização das proteínas durante a diferenciação. A proteína TcRBP40 será nocauteada e seu efeito avaliado no parasita. No entanto, isso seria desafiador para a proteína TcRBP7, uma vez que seu gene codificante apresenta múltiplas cópias no genoma. Vetores de superexpressão no parasita serão utilizados para verificar os efeitos do excesso das proteínas no parasita.

## REFERÊNCIAS

- ABUIN, G. et al. Expression of trans-Sialidase and 85-kDa Glycoprotein Genes in *Trypanosoma cruzi* Is Differentially Regulated at the Post-transcriptional Level by Labile Protein Factors. **The Journal of Biological Chemistry**, v. 274, n. 19, p. 13041–13047, 1999.
- ALBERTS, B. et al. **Biologia Molecular da Célula**. 5. ed. [s.l.] Artmed, 2010.
- ALVES, L. R. et al. mRNA Localization Mechanisms in *Trypanosoma cruzi*. **PLoS One**, v. 8, n. 12, p. e81375, jan. 2013.
- ANDRADE, L. O.; ANDREWS, N. W. Lysosomal fusion is essential for the retention of *Trypanosoma cruzi* inside host cells. **The Journal of Experimental Medicine**, v. 200, n. 9, p. 1135–43, 1 nov. 2004.
- ARAÚJO, P. R. et al. Development of a dual reporter system to identify regulatory cis-acting elements in untranslated regions of *Trypanosoma cruzi* mRNAs. **Parasitology International**, v. 60, n. 2, p. 161–169, 2011.
- AUWETER, S. D.; OBERSTRASS, F. C.; ALLAIN, F. H.-T. Sequence-specific binding of single-stranded RNA: is there a code for recognition? **Nucleic Acids Research**, v. 34, n. 17, p. 4943–4959, 2006.
- AVILA, A. R. et al. Stage-specific gene expression during *Trypanosoma cruzi* metacyclogenesis. **Genetics and Molecular Research**, v. 2, n. 1, p. 159–68, jan. 2003.
- BAILEY, T. L.; ELKAN, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. **Proceedings / International Conference on Intelligent Systems for Molecular Biology**, v. 2, p. 28–36, jan. 1994.
- BALAKIREV, E. S.; AYALA, F. J. Pseudogenes: are they “junk” or functional DNA? **Annual Review of Genetics**, v. 37, p. 123–151, 2003.
- BATISTA, C. M. et al. mAb CZP-315.D9: An Antirecombinant Cruzipain Monoclonal Antibody That Specifically Labels the Reservosomes of *Trypanosoma cruzi* Epimastigotes. **BioMed Research International**, v. 2014, p. 714749, 2014.
- BATISTA, M. et al. A high-throughput cloning system for reverse genetics in *Trypanosoma cruzi*. **BMC microbiology**, v. 10, n. 1, p. 259, jan. 2010.
- BRIONES, M. R. et al. The evolution of two *Trypanosoma cruzi* subgroups inferred from rRNA genes can be correlated with the interchange of American mammalian faunas in the Cenozoic and has implications to pathogenicity and host specificity. **Molecular and biochemical parasitology**, v. 104, n. 2, p. 219–32, 30 nov. 1999.
- BUTTER, F. et al. Unbiased RNA-protein interaction screen by quantitative proteomics. **Proceedings of the National Academy of Sciences of the United States of America**, v. 106, n. 26, p. 10626–10631, 2009.
- CAMPBELL, D. A.; THOMAS, S.; STURM, N. R. Transcription in kinetoplastid protozoa: why be normal? **Microbes and Infection**, v. 5, n. 13, p. 1231–1240, nov. 2003.
- CAMPOS, P. C. et al. Sequences involved in mRNA processing in *Trypanosoma cruzi*. **International Journal for Parasitology**, v. 38, n. 12, p. 1383–9, out. 2008.
- CASSOLA, A.; FRASCH, A. C. An RNA recognition motif mediates the nucleocytoplasmic transport of a trypanosome RNA-binding protein. **The Journal of Biological Chemistry**, v. 284, n. 50, p. 35015–28, 11 dez. 2009.
- CECH, T. R.; STEITZ, J. A. The noncoding RNA revolution-trashing old rules to forge new ones. **Cell**, v. 157, n. 1, p. 77–94, 27 mar. 2014.
- CHAGAS, C. Nova tripanozomíase humana. **Memórias do Instituto Oswaldo Cruz**,



v. 1, n. 2, p. 159–218, 1909.

CHAGAS, E. Summula dos conhecimentos actuaes sobre a Trypanosomiasis americana. **Memorias do Instituto Oswaldo Cruz**, v. 30, n. 3, p. 378–416, 1935.

CHEN, Y.; VARANI, G. Engineering RNA-binding proteins for biology. **The FEBS Journal**, v. 280, n. 16, p. 3734–54, ago. 2013.

CLAYTON, C. E. Life without transcriptional control? From fly to man and back again. **EMBO Journal**, v. 21, n. 8, p. 1881–1888, 2002.

CLAYTON, C.; SHAPIRA, M. Post-transcriptional regulation of gene expression in trypanosomes and leishmanias. **Molecular & Biochemical Parasitology**, v. 156, p. 93–101, 2007.

CLÉRY, A.; BLATTER, M.; ALLAIN, F. H.-T. RNA recognition motifs: boring? Not quite. **Current opinion in structural biology**, v. 18, n. 3, p. 290–8, jun. 2008.

CONANT, G. C.; WOLFE, K. H. Turning a hobby into a job: how duplicated genes find new functions. **Nature reviews. Genetics**, v. 9, n. 12, p. 938–50, dez. 2008.

CUNHA-E-SILVA, N. et al. Isolation and characterization of a reservosome fraction from *Trypanosoma cruzi*. **FEMS Microbiology Letters**, v. 214, p. 7–12, 2002.

CUROTTO DE LAFAILLE, M. A.; LABAN, A.; WIRTH, D. F. Gene expression in *Leishmania*: analysis of essential 5' DNA sequences. **Proceedings of the National Academy of Sciences of the United States of America**, v. 89, n. 7, p. 2703–7, 1 abr. 1992.

D'ORSO, I.; FRASCH, A. C. Functionally different AU- and G-rich cis-elements confer developmentally regulated mRNA stability in *Trypanosoma cruzi* by interaction with specific RNA-binding proteins. **The Journal of Biological Chemistry**, v. 276, n. 19, p. 15783–93, 11 maio 2001a.

D'ORSO, I.; FRASCH, A. C. TcUBP-1, a developmentally regulated U-rich RNA-binding protein involved in selective mRNA destabilization in trypanosomes. **The Journal of Biological Chemistry**, v. 276, n. 37, p. 34801–9, 14 set. 2001b.

D'ORSO, I.; FRASCH, A. C. C. TcUBP-1, an mRNA destabilizing factor from trypanosomes, homodimerizes and interacts with novel AU-rich element- and Poly(A)-binding proteins forming a ribonucleoprotein complex. **The Journal of biological chemistry**, v. 277, n. 52, p. 50520–8, 27 dez. 2002.

DAUBNER, G. M.; CLÉRY, A.; ALLAIN, F. H.-T. RRM-RNA recognition: NMR or crystallography...and new findings. **Current opinion in structural biology**, v. 23, n. 1, p. 100–8, fev. 2013.

DE GODOY, L. M. F. et al. Quantitative proteomics of *Trypanosoma cruzi* during metacyclogenesis. **Proteomics**, v. 12, n. 17, p. 2694–2703, 2012.

DE SOUZA, W. Basic cell biology of *Trypanosoma cruzi*. **Current Pharmaceutical Design**, v. 8, n. 4, p. 269–85, jan. 2002.

DE SOUZA, W. Structural organization of *Trypanosoma cruzi*. **Memórias do Instituto Oswaldo Cruz**, v. 104, n. I, p. 89–100, 2009.

DEKOSTER, G. T.; DELANEY, K. J.; HALL, K. B. A Compare-and-Contrast NMR Dynamics Study of Two Related RRMs: U1A and SNF. **Biophysical Journal**, v. 107, n. 1, p. 208–19, 1 jul. 2014.

DONG, S. et al. Specific and modular binding code for cytosine recognition in Pumilio/FBF (PUF) RNA-binding domains. **Journal of Biological Chemistry**, v. 286, n. 30, p. 26732–26742, 2011.

EGER, I.; SOARES, M. J. Endocytosis in *Trypanosoma cruzi* (Euglenozoa: Kinetoplastea) epimastigotes: visualization of ingested transferrin-gold nanoparticle

- complexes by confocal laser microscopy. **Journal of Microbiological Methods**, v. 91, n. 1, p. 101–105, out. 2012.
- EL-SAYED, N. M. et al. The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. **Science**, v. 309, n. 409, p. 409–15, 15 jul. 2005.
- ELIAS, M. C. Q. B. et al. Transcription rate modulation through the *Trypanosoma cruzi* life cycle occurs in parallel with changes in nuclear organisation. **Molecular and Biochemical Parasitology**, v. 112, n. 1, p. 79–90, 15 jan. 2001.
- EWING, B. et al. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. **Genome Research**, v. 8, n. 3, p. 175–85, mar. 1998.
- FARINA, K. L. et al. Two ZBP1 KH domains facilitate beta-actin mRNA localization, granule formation, and cytoskeletal attachment. **The Journal of cell biology**, v. 160, n. 1, p. 77–87, 6 jan. 2003.
- FERREIRA, L. R. P. et al. Active transcription and ultrastructural changes during *Trypanosoma cruzi* metacyclogenesis. **Anais da Academia Brasileira de Ciências**, v. 80, n. 1, p. 157–66, mar. 2008.
- FIRE, A. et al. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. **Nature**, v. 391, n. February, p. 806–811, 1998.
- GARCIA-SILVA, M. R. et al. A population of tRNA-derived small RNAs is actively produced in *Trypanosoma cruzi* and recruited to specific cytoplasmic granules. **Molecular & Biochemical Parasitology**, v. 171, n. 2, p. 64–73, 2010.
- GARCIA-SILVA, M. R. et al. Extracellular vesicles shed by *Trypanosoma cruzi* are linked to small RNA pathways, life cycle regulation, and susceptibility to infection of mammalian cells. **Parasitology Research**, v. 113, n. 1, p. 285–304, jan. 2014a.
- GARCIA-SILVA, M. R. et al. Gene expression changes induced by *Trypanosoma cruzi* shed microvesicles in mammalian host cells: relevance of tRNA-derived halves. **BioMed research international**, v. 2014, p. 305239, jan. 2014b.
- GARCIA-SILVA, M. R. et al. A particular set of small non-coding RNAs is bound to the distinctive Argonaute protein of *Trypanosoma cruzi*: Insights from RNA-interference deficient organisms. **Gene**, v. 538, n. 2, p. 379–84, 1 abr. 2014c.
- GARCIA, E.; GONZALEZ, M.; AZAMBUJA, P. Biological factors involving *Trypanosoma cruzi* life cycle in the invertebrate vector, *Rhodnius prolixus*. **Memórias do Instituto Oswaldo Cruz**, v. 94, n. 1, p. 213–6, jan. 1999.
- GAUDENZI, J. G. DE; FRASCH, A. C.; CLAYTON, C. RNA-binding domain proteins in kinetoplastids: a comparative analysis. **Eukaryotic cell**, v. 4, n. 12, p. 2106–2114, 2005.
- GERBER, P.; HERSCHLAG, D.; BROWN, P. O. Extensive Association of Functionally and Cytotopically Related mRNAs with Puf Family RNA-Binding Proteins in Yeast. **PLoS Biology**, v. 2, n. 3, 2004.
- GILINGER, G.; BELLOFATTO, V. Trypanosome spliced leader RNA genes contain the first identified RNA polymerase II gene promoter in these organisms. **Nucleic Acids Research**, v. 29, n. 7, p. 1556–64, 1 abr. 2001.
- GIRARD-DIAS, W. et al. On the ultrastructural organization of *Trypanosoma cruzi* using cryopreparation methods and electron tomography. **Histochemistry and Cell Biology**, v. 138, n. 6, p. 821–831, 2012.
- GLISOVIC, T. et al. RNA-binding proteins and post-transcriptional gene regulation. **FEBS letters**, v. 582, n. 14, p. 1977–86, 18 jun. 2008.
- GUERRA-SLOMPO, E. P. et al. Molecular characterization of the *Trypanosoma cruzi* specific RNA binding protein TcRBP40 and its associated mRNAs.

- Biochemical and Biophysical Research Communications**, v. 420, n. 2, p. 302–307, 2012.
- HAILE, S.; PAPADOPOULOU, B. Developmental regulation of gene expression in trypanosomatid parasitic protozoa. **Current Opinion in Microbiology**, v. 10, p. 569–577, 2007.
- HE, X.; ZHANG, J. Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. **Genetics**, v. 169, n. 2, p. 1157–64, fev. 2005.
- HOGAN, D. J. et al. Diverse RNA-Binding Proteins Interact with Functionally Related Sets of RNAs, Suggesting an Extensive Regulatory System. **PLoS Biology**, v. 6, n. 10, p. e255, 2008.
- HOLETZ, B. et al. Protein and mRNA content of TcDHH1-containing mRNPs in *Trypanosoma cruzi*. **FEBS Journal**, v. 277, p. 3415–3426, 2010.
- HUNTLEY, R. P. et al. A method for increasing expressivity of Gene Ontology annotations using a compositional approach. **BMC bioinformatics**, v. 15, n. 1, p. 155, jan. 2014.
- HURLEY, J. H.; EMR, S. D. The ESCRT complexes: structure and mechanism of a membrane-trafficking network. **Annual review of biophysics and biomolecular structure**, v. 35, p. 277–98, jan. 2006.
- IRION, U.; ST JOHNSTON, D. bicoid RNA localization requires specific binding of an endosomal sorting complex. **Nature**, v. 445, n. 7127, p. 554–8, 1 fev. 2007.
- JENSEN, B. C. et al. Widespread variation in transcript abundance within and across developmental stages of *Trypanosoma brucei*. **BMC Genomics**, v. 10, p. 482, jan. 2009.
- KAFASLA, P. et al. Defining the roles and interactions of PTB. **Biochemical Society transactions**, v. 40, n. 4, p. 815–20, ago. 2012.
- KAFKOVÁ, L. et al. Functional characterization of two paralogs that are novel RNA binding proteins influencing mitochondrial transcripts of *Trypanosoma brucei*. **RNA**, v. 18, n. 10, p. 1846–61, out. 2012.
- KEENE, J. D. Ribonucleoprotein infrastructure regulating the flow of genetic information between the genome and the proteome. **Proceedings of the National Academy of Sciences of the United States of America**, v. 98, n. 13, p. 7018–7024, 2001.
- KEENE, J. D. RNA regulons: coordination of post-transcriptional events. **Nature Reviews - Genetics**, v. 8, n. July, p. 533–543, 2007.
- KEENE, J. D.; TENENBAUM, S. A. Eukaryotic mRNPs may represent posttranscriptional operons. **Molecular Cell**, v. 9, n. 6, p. 1161–7, 1 jun. 2002.
- KISLAUSKIS, E. H.; ZHU, X.; SINGER, R. H. Sequences responsible for intracellular localization of beta-actin messenger RNA also affect cell phenotype. **The Journal of cell biology**, v. 127, n. 2, p. 441–51, out. 1994.
- KITANO, H. Systems biology: a brief overview. **Science**, v. 295, n. 5560, p. 1662–4, 1 mar. 2002.
- KOONIN, E. V. Darwinian evolution in the light of genomics. **Nucleic Acids Research**, v. 37, n. 4, p. 1011–34, mar. 2009.
- KOONIN, E. V. Does the central dogma still stand? **Biology Direct**, v. 7, n. 27, p. 1–7, jan. 2012.
- KRAMER, S. Developmental regulation of gene expression in the absence of transcriptional control: the case of kinetoplastids. **Molecular and biochemical**

- parasitology**, v. 181, n. 2, p. 61–72, fev. 2012.
- LEBOWITZ, J. H. et al. Coupling of poly(A) site selection and trans-splicing in *Leishmania*. **Genes & Development**, v. 7, n. 6, p. 996–1007, 1 jun. 1993.
- LIANG, X. et al. trans and cis Splicing in Trypanosomatids: Mechanism, Factors, and Regulation. **Eukaryotic Cell**, v. 2, n. 5, p. 830–840, 2003.
- LONG, M. et al. The origin of new genes: glimpses from the young and old. **Nature Reviews - Genetics**, v. 4, n. 11, p. 865–75, nov. 2003.
- LUNDE, B. M.; MOORE, C.; VARANI, G. RNA-binding proteins: modular design for efficient function. **Nature Reviews - Molecular Cell Biology**, v. 8, n. June, p. 479–490, 2007.
- LYNCH, M.; FORCE, A. The probability of duplicate gene preservation by subfunctionalization. **Genetics**, v. 154, n. 1, p. 459–73, jan. 2000.
- LYNCH, M.; KATJU, V. The altered evolutionary trajectories of gene duplicates. **Trends in Genetics**, v. 20, n. 11, p. 544–9, nov. 2004.
- MAIR, G. et al. A new twist in trypanosome RNA metabolism: cis-splicing of pre-mRNA. **RNA**, v. 6, n. 2, p. 163–9, fev. 2000.
- MANOEL-CAETANO, F. DA S.; SILVA, A. E. Implications of genetic variability of *Trypanosoma cruzi* for the pathogenesis of Chagas disease. **Cadernos de Saúde Pública**, v. 23, n. 10, p. 2263–74, out. 2007.
- MAO, Y.; NAJAFABADI, H. S.; SALAVATI, R. Genome-wide computational identification of functional RNA elements in *Trypanosoma brucei*. **BMC Genomics**, v. 10, n. 355, p. 355, jan. 2009.
- MARTIN, K. C.; EPHRUSSI, A. mRNA localization: gene expression in the spatial dimension. **Cell**, v. 136, n. 4, p. 719–30, 20 fev. 2009.
- MARTÍNEZ-CALVILLO, S. et al. Transcription of *Leishmania major* Friedlin chromosome 1 initiates in both directions within a single region. **Molecular Cell**, v. 11, n. 5, p. 1291–9, maio 2003.
- MARTÍNEZ-CALVILLO, S. et al. Gene expression in trypanosomatid parasites. **Journal of Biomedicine & Biotechnology**, v. 2010, p. 525241, jan. 2010.
- MASLOV, D. A.; PODLIPAEV, S. A.; LUKE, J. Phylogeny of the Kinetoplastida: Taxonomic Problems and Insights into the Evolution of Parasitism. **Memorias do Instituto Oswaldo Cruz**, v. 96, n. 3, p. 397–402, 2001.
- MATTICK, J. S. The genetic signatures of noncoding RNAs. **PLoS genetics**, v. 5, n. 4, p. e1000459, abr. 2009.
- MCANDREW, M. et al. Testing promoter activity in the trypanosome genome: isolation of a metacyclic-type VSG promoter, and unexpected insights into RNA polymerase II transcription. **Experimental parasitology**, v. 90, n. 1, p. 65–76, set. 1998.
- MIGNONE, F. et al. Untranslated regions of mRNAs. **Genome Biology**, v. 3, n. 3, p. REVIEWS0004, jan. 2002.
- MONTGOMERY, M. K.; XU, S.; FIRE, A. RNA as a target of double-stranded RNA-mediated genetic interference in *Caenorhabditis elegans*. **Proceedings of the National Academy of Sciences of the United States of America**, v. 95, n. 26, p. 15502–7, 22 dez. 1998.
- MONY, B. M. et al. Genome-wide dissection of the quorum sensing signalling pathway in *Trypanosoma brucei*. **Nature**, v. 505, n. 7485, p. 681–5, 30 jan. 2014.
- MORRIS, J. C. et al. Replication of kinetoplast DNA: an update for the new millennium. **International Journal for Parasitology**, v. 31, n. 5-6, p. 453–8, 1 maio

2001.

NOÉ, G.; GAUDENZI, J. G. DE; FRASCH, A. C. Functionally related transcripts have common RNA motifs for specific RNA-binding proteins in trypanosomes. **BMC Molecular Biology**, v. 9, n. 107, p. 1–19, 2008.

NOWAK, M. A et al. Evolution of genetic redundancy. **Nature**, v. 388, n. 6638, p. 167–71, 10 jul. 1997.

PARK, C. et al. Differential requirements for mRNA folding partially explain why highly expressed proteins evolve slowly. **Proceedings of the National Academy of Sciences of the United States of America**, v. 110, n. 8, p. E678–86, 19 fev. 2013.

PATTHY, L. **Protein Evolution**. 2. ed. [s.l.] Wiley-Blackwell, 2007.

PÉREZ-DÍAZ, L. et al. Trypanosoma cruzi: molecular characterization of an RNA binding protein differentially expressed in the parasite life cycle. **Experimental parasitology**, v. 117, n. 1, p. 99–105, set. 2007.

PÉREZ-DÍAZ, L. et al. The overexpression of the trypanosomatid-exclusive TcRBP19 RNA-binding protein affects cellular infection by Trypanosoma cruzi. **Memórias do Instituto Oswaldo Cruz**, v. 107, n. 8, p. 1076–9, dez. 2012.

PÉREZ-DÍAZ, L. et al. Evidence for a negative feedback control mediated by the 3' untranslated region assuring the low expression level of the RNA binding protein TcRBP19 in T. cruzi epimastigotes. **Biochemical and biophysical research communications**, v. 436, n. 2, p. 295–9, 28 jun. 2013.

PHILIPPE, H. et al. Early-branching or fast-evolving eukaryotes? An answer based on slowly evolving positions. **Proceedings of the Royal Society - Biological Sciences**, v. 267, n. 1449, p. 1213–1221, 2000.

REIGNAULT, L. C. et al. Structures containing galectin-3 are recruited to the parasitophorous vacuole containing Trypanosoma cruzi in mouse peritoneal macrophages. **Parasitology Research**, v. 113, n. 6, p. 2323–2333, 2014.

ROSS, A. F. et al. Characterization of a beta-actin mRNA zipcode-binding protein. **Molecular and Cellular Biology**, v. 17, n. 4, p. 2158–65, abr. 1997.

RUSTEN, T. E.; STENMARK, H. Developmental biology: moonlighting at the pole. **Nature**, v. 445, n. 7127, p. 497–9, 1 fev. 2007.

SABIN, L. R.; DELÁS, M. J.; HANNON, G. J. Dogma derailed: the many influences of RNA on the genome. **Molecular Cell**, v. 49, n. 5, p. 783–94, 7 mar. 2013.

SAMBROOK, J.; RUSSEL, D. W. **Molecular Cloning: A Laboratory Manual**. 3. ed. [s.l.] Cold Spring Harbor Laboratory Press, 2001.

SANCHEZ-DIAZ, P.; PENALVA, L. O. F. Post-transcription meets post-genomic: the saga of RNA binding proteins in a new era. **RNA Biology**, v. 3, n. 3, p. 101–109, 2006.

SANT'ANNA, C. et al. All Trypanosoma cruzi developmental forms present lysosome-related organelles. **Histochemistry and Cell Biology**, v. 130, n. 6, p. 1187–1198, 2008a.

SANT'ANNA, C. et al. New Insights Into the Morphology of Trypanosoma cruzi Reservosome. **Microscopy Research and Technique**, v. 71, p. 599–605, 2008b.

SANT'ANNA, C.; DE SOUZA, W.; CUNHA-E-SILVA, N. Biogenesis of the Reservosomes of Trypanosoma cruzi. **Microscopy and Microanalysis**, v. 10, p. 637–646, 2004.

SCHERRER, T. et al. A Screen for RNA-Binding Proteins in Yeast Indicates Dual Functions for Many Enzymes. **PLoS One**, v. 5, n. 11, p. e15499, 2010.

SCOFIELD, D. G.; LYNCH, M. Evolutionary diversification of the Sm family of

- RNA-associated proteins. **Molecular biology and evolution**, v. 25, n. 11, p. 2255–67, nov. 2008.
- SLAGSVOLD, T. et al. Endosomal and non-endosomal functions of ESCRT proteins. **Trends in Cell Biology**, v. 16, n. 6, p. 317–26, jun. 2006.
- SOARES, M. J.; DE SOUZA, W. Cytoplasmic organelles of trypanosomatids: a cytochemical and stereological study. **Journal of Submicroscopic Cythology and Pathology**, v. 20, n. 2, p. 349–361, 1988.
- SOARES, M. J.; DE SOUZA, W. Endocytosis of gold-labeled proteins and LDL by *Trypanosoma cruzi*. **Parasitology Research**, v. 77, n. 6, p. 461–468, 1991.
- SOARES, M. J.; SOUTO-PADRÓN, T.; DE SOUZA, W. Identification of a large pre-lysosomal compartment in the pathogenic protozoon *Trypanosoma cruzi*. **Journal of Cell Science**, v. 102, p. 157–167, maio 1992.
- STUART, K. et al. Kinetoplastids: related protozoan pathogens, different diseases. **The Journal of Clinical Investigation**, v. 118, n. 4, p. 1301–10, abr. 2008.
- STUDIER, F. W. et al. Use of T7 RNA polymerase to direct expression of cloned genes. **Methods in Enzymology**, v. 185, p. 60–89, jan. 1990.
- TAMURA, K. et al. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. **Molecular Biology and Evolution**, v. 24, n. 8, p. 1596–9, ago. 2007.
- TAN, T. H. P. et al. tRNAs in *Trypanosoma brucei*: genomic organization, expression, and mitochondrial import. **Molecular and Cellular Biology**, v. 22, n. 11, p. 3707–17, jun. 2002.
- TENENBAUM, S. A. et al. Identifying mRNA subsets in messenger ribonucleoprotein complexes by using cDNA arrays. **Proceedings of the National Academy of Sciences of the United States of America**, v. 97, n. 26, p. 14085–90, 19 dez. 2000.
- TOWNLEY-TILSON, W. H. D. et al. Genome-wide analysis of mRNAs bound to the histone stem-loop binding protein. **RNA**, v. 12, p. 1853–1867, 2006.
- TYLER, K. M.; ENGMAN, D. M. The life cycle of *Trypanosoma cruzi* revisited. **International journal for parasitology**, v. 31, n. 5-6, p. 472–481, 2001.
- VANHAMME, L.; PAYS, E. Control of Gene Expression in Trypanosomes. **Microbiological Reviews**, v. 59, n. 2, p. 223–240, 1995.
- WEATHERLY, D. B.; BOEHLKE, C.; TARLETON, R. L. Chromosome level assembly of the hybrid *Trypanosoma cruzi* genome. **BMC Genomics**, v. 10, n. 255, jan. 2009.
- ZHANG, J. Evolution by gene duplication: an update. **Trends in Ecology & Evolution**, v. 18, n. 6, p. 292–298, jun. 2003.
- ZHAO, L. et al. Origin and spread of de novo genes in *Drosophila melanogaster* populations. **Science**, v. 343, n. 6172, p. 769–72, 14 fev. 2014.
- ZINGALES, B. et al. *Trypanosoma cruzi* genome project: biological characteristics and molecular typing of clone CL Brener. **Acta Tropica**, v. 68, n. 2, p. 159–73, nov. 1997.
- ZINGALES, B. et al. A new consensus for *Trypanosoma cruzi* intraspecific nomenclature: second revision meeting recommends TcI to TcVI. **Memórias do Instituto Oswaldo Cruz**, v. 104, n. 7, p. 1051–4, nov. 2009.
- ZINOVIEV, A. et al. Gene duplication in trypanosomatids – Two DED1 paralogs are functionally redundant and differentially expressed during the life cycle. **Molecular & Biochemical Parasitology**, p. 1–10, 2012.

## **ANEXO**

As duas tabelas abaixo mostram os alvos das proteínas TcRBP7 (1ª Tabela) e TcRBP40 (2ª Tabela), de acordo com a seleção pelos 5 filtros (Seção 4.5.4.4). Cada alvo contém os valores de para os 3 experimentos (2 RBPs e o controle), para fins de comparação.

Identificação Alvos de TcRBP7	Anotação do gene Alvos de TcRBP7	Filtro 1: Número de réplicas enriquecidas na fração ligada em TcRBP7	Filtro 1: Número de réplicas enriquecidas na fração ligada em TcRBP40	Filtro 2: Média de leituras da fração ligada em TcRBP7	Filtro 2: Média de leituras da fração ligada em TcRBP40	Filtro 3: Média FC em Controle	Filtro 3: Média FC em TcRBP7	Filtro 3: Média FC em TcRBP40	Filtro 4: Número de réplicas com FC $\geq 2$ em Controle	Filtro 4: Número de réplicas com FC $\geq 2$ em TcRBP7	Filtro 4: Número de réplicas com FC $\geq 2$ em TcRBP40	Filtro 5: $\Delta$ FC Médio (FC TcRBP7 - FC Controle)	Filtro 5: $\Delta$ FC Médio (FC TcRBP40 - FC Controle)	FC RBP7 / FC RBP40	FC RBP40 / FC RBP7
TcCLB.408345.10	hypothetical protein, conserved	3	2	6,3	2,0	0,000	2,383	2,292	0	2	1	2,383	2,292	1,040	0,962
TcCLB.408437.29	hypothetical protein	3	0	12,7	1,0	-1,091	5,026	0,000	0	3	0	6,116	1,091	5,607	0,178
TcCLB.418221.4	hypothetical protein, conserved	3	2	15,0	3,0	2,761	6,813	3,491	2	2	1	4,053	0,730	5,551	0,180
TcCLB.458759.14	hypothetical protein, conserved receptor-type adenylate cyclase, putative	3	2	11,0	2,3	-2,681	6,267	4,032	1	2	1	8,948	6,714	1,333	0,750
TcCLB.460757.10	hypothetical protein, conserved	3	3	8,0	6,0	-3,462	2,988	6,017	0	2	3	6,450	9,480	0,680	1,470
TcCLB.484299.10	trypanothione reductase, putative	3	3	13,3	5,0	2,372	8,533	1,949	1	2	1	6,161	-0,423	-14,560	-0,069
TcCLB.485889.4	hypothetical protein, conserved	3	1	8,7	6,3	-2,553	5,281	4,874	0	3	1	7,833	7,427	1,055	0,948
TcCLB.503395.30	hypothetical protein, conserved	3	1	5,3	4,3	-2,944	2,124	4,041	1	2	1	5,067	6,984	0,726	1,378
TcCLB.503561.20	hypothetical protein, conserved	3	3	6,3	7,0	1,920	5,058	3,418	1	2	2	3,137	1,497	2,095	0,477
TcCLB.503565.10	hypothetical protein	3	1	5,7	0,7	-2,487	3,790	7,452	0	2	1	6,277	9,939	0,632	1,583
TcCLB.503581.30	hypothetical protein, conserved proteasome alpha 7 subunit, putative	3	1	25,3	11,3	-1,487	3,669	1,916	0	3	0	5,156	3,403	1,515	0,660
TcCLB.503613.20	hypothetical protein, conserved	3	3	12,3	15,7	0,000	3,068	4,053	0	2	3	3,068	4,053	0,757	1,321
TcCLB.503657.9	hypothetical protein, conserved	3	3	49,0	31,3	-1,142	2,883	2,587	0	2	2	4,024	3,729	1,079	0,927
TcCLB.503697.130	hypothetical protein, conserved	3	2	14,0	6,7	2,307	5,082	4,644	1	3	1	2,775	2,337	1,188	0,842
TcCLB.503773.50	nucleolar protein, putative	3	2	5,7	6,0	2,428	5,334	1,980	2	3	1	2,906	-0,448	-6,485	-0,154
TcCLB.503789.70	hypothetical protein, conserved	3	2	23,3	11,0	-1,286	2,102	2,154	0	2	1	3,388	3,440	0,985	1,015
TcCLB.503801.50	hypothetical protein, conserved phosphatidylinositol-4-phosphate 5-kinase, putative	3	0	9,0	4,7	-2,042	2,215	0,000	0	2	0	4,258	2,042	2,085	0,480
TcCLB.503891.30	NADH-ubiquinone oxidoreductase 20 kDa subunit, mitochondrial precursor, putative	3	0	9,7	4,7	-2,874	2,039	0,000	0	2	0	4,913	2,874	1,710	0,585
TcCLB.503893.80	hypothetical protein	3	2	13,3	6,7	0,000	4,832	2,188	1	2	1	4,832	2,188	2,209	0,453
TcCLB.503909.90	protein kinase, putative translation initiation factor IF-2, putative	3	1	6,7	4,0	-1,300	4,211	3,481	0	3	1	5,510	4,781	1,153	0,868
TcCLB.504035.90	hypothetical protein, conserved	3	1	9,0	3,3	2,989	5,550	1,242	2	2	0	2,561	-1,747	-1,466	-0,682
TcCLB.504045.80	ribosomal RNA processing protein	3	3	187,3	95,0	2,140	5,310	5,009	2	2	3	3,170	2,869	1,105	0,905



6, putative,exosome complex  
exonuclease, putative (RRP6)

TcCLB.504109.10	hypothetical protein, conserved	3	3	117,0	64,0	-2,007	3,625	2,571	0	3	2	5,632	4,578	1,230	0,813
TcCLB.504255.30	hypothetical protein, conserved	3	1	5,0	2,0	0,000	3,307	2,329	0	3	1	3,307	2,329	1,420	0,704
TcCLB.504625.10	trans-splicing factor, putative polynucleotide kinase 3'-	3	3	13,0	10,3	1,336	3,413	3,606	0	2	1	2,077	2,270	0,915	1,093
TcCLB.505807.190	phosphatase, putative	3	1	5,3	4,3	-2,140	3,874	0	0	3	1	6,014	0	0,396	2,524
TcCLB.505919.40	protein kinase, putative	3	3	45,0	41,7	-3,284	3,021	2,151	1	2	2	6,305	5,435	1,160	0,862
TcCLB.505965.60	hypothetical protein, conserved	3	1	5,3	0,3	0,000	3,448	1,393	1	2	0	3,448	1,393	2,476	0,404
TcCLB.506135.50		3	3	24,7	23,7	2,055	4,055	2,845	2	2	3	2,000	0,790	2,533	0,395
TcCLB.506145.10	hypothetical protein, conserved	3	2	36,0	26,3	-1,394	2,245	1,432	0	2	0	3,639	2,826	1,288	0,777
TcCLB.506173.150	hypothetical protein	3	2	9,3	7,3	2,461	4,779	3,386	1	2	1	2,318	0,925	2,506	0,399
TcCLB.506251.110	hypothetical protein, conserved	3	1	8,7	3,3	0,000	7,351	3,105	1	3	1	7,351	3,105	2,367	0,422
TcCLB.506327.99	serine/threonine protein kinase, putative	3	1	5,0	0,7	0,000	4,377	2,785	1	3	1	4,377	2,785	1,572	0,636
TcCLB.506333.10	plasma-membrane proton-efflux P-type ATPase, putative	3	3	31,0	9,3	3,742	11,652	5,029	2	3	2	7,910	1,287	6,148	0,163
TcCLB.506335.100	mucin-associated surface protein (MASP), putative	3	3	8,0	6,7	0,000	3,278	2,639	0	2	2	3,278	2,639	1,242	0,805
TcCLB.506351.50	hypothetical protein	3	2	5,3	3,0	1,780	5,561	1,628	1	2	0	3,782	-0,152	-24,889	-0,040
TcCLB.506355.60	small G-protein, putative	3	3	20,7	5,0	-3,482	2,436	2,760	0	2	1	5,918	6,242	0,948	1,055
TcCLB.506357.190	hypothetical protein, conserved	3	3	10,3	13,0	2,112	5,924	3,750	1	2	2	3,813	1,638	2,327	0,430
TcCLB.506357.80	hypothetical protein, conserved	3	2	5,7	3,0	0,000	2,924	1,727	0	2	0	2,924	1,727	1,693	0,591
TcCLB.506407.20		3	3	17,3	12,3	1,235	3,448	2,065	0	2	2	2,213	0,830	2,665	0,375
TcCLB.506409.230	hypothetical protein, conserved	3	3	42,7	31,7	9,470	22,674	0	2	3	3	4	6,170	2,140	0,467
TcCLB.506477.60	protein kinase, putative	3	3	11,0	10,7	-2,187	4,354	2,342	0	2	1	6,540	4,528	1,444	0,692
TcCLB.506479.74	FG-GAP repeat protein, putative	3	2	25,3	14,3	-1,647	4,394	2,868	1	2	1	6,041	4,514	1,338	0,747
TcCLB.506493.40	hypothetical protein, conserved	3	0	6,3	2,7	-1,449	2,183	0,000	0	2	0	3,632	1,449	2,506	0,399
TcCLB.506525.90	hypothetical protein, conserved	3	3	8,0	2,7	0,000	2,499	2,160	0	3	2	2,499	2,160	1,157	0,864
TcCLB.506529.50	arginine N-methyltransferase, putative	3	3	83,7	27,3	3,712	5,798	4,002	3	3	3	2,085	0,290	7,194	0,139
TcCLB.506575.40	hypothetical protein	3	1	15,7	8,0	-1,631	2,052	1,102	0	2	0	3,683	2,733	1,348	0,742
TcCLB.506577.47	hypothetical protein, conserved	3	1	10,3	3,0	0,000	4,000	2,293	1	2	1	4,000	2,293	1,745	0,573
TcCLB.506579.100		3	2	11,3	6,3	0,000	2,233	1,520	0	2	0	2,233	1,520	1,469	0,681
TcCLB.506629.20	hypothetical protein, conserved	3	1	6,3	1,3	-2,327	2,774	1,490	0	2	0	5,101	3,817	1,336	0,748
TcCLB.506635.130	hypothetical protein, conserved	3	1	10,7	3,3	-3,408	3,087	1,397	0	2	0	6,495	4,805	1,352	0,740
TcCLB.506657.79	50S ribosomal protein L16, putative	3	2	16,7	6,0	1,249	4,346	1,526	0	2	0	3,097	0,277	11,183	0,089
TcCLB.506661.10	fatty acid elongase, putative	3	2	8,3	1,7	0,000	3,680	2,559	0	2	1	3,680	2,559	1,438	0,696

												18,70			
TcCLB.506661.160	hypothetical protein, conserved	3	3	70,7	24,3	-2,239	16,463	6,260	0	2	2	2	8,499	2,201	0,454
TcCLB.506691.80	hypothetical protein, conserved	3	2	17,7	6,0	5,714	11,259	4,083	2	3	2	5,546	-1,631	-3,401	-0,294
TcCLB.506693.10	hypothetical protein	3	3	10,7	23,3	0,000	6,641	7,431	0	3	2	6,641	7,431	0,894	1,119
TcCLB.506705.10	hypothetical protein	3	1	13,3	7,3	-3,521	2,790	1,393	0	3	0	6,311	4,914	1,284	0,779
TcCLB.506753.230	hypothetical protein, conserved	3	3	11,7	3,3	2,146	6,735	2,064	1	2	1	4,589	-0,082	-56,290	-0,018
TcCLB.506775.140	hypothetical protein, conserved	3	2	14,3	5,7	0,000	2,823	5,523	0	3	1	2,823	5,523	0,511	1,956
	protein kinase, putative, serine/threonine protein kinase														
TcCLB.506775.190	kinase, putative	3	2	8,7	4,7	-1,462	4,144	2,584	0	3	1	5,606	4,047	1,385	0,722
TcCLB.506777.24	hypothetical protein, conserved	3	1	8,0	3,7	-1,520	6,523	1,597	1	3	0	8,043	3,117	2,581	0,388
TcCLB.506793.30	kinesin, putative	3	2	21,3	12,7	-2,511	2,108	5,890	1	2	1	4,619	8,401	0,550	1,819
TcCLB.506811.60	hypothetical protein	3	1	30,3	7,0	2,880	6,000	3,992	2	3	1	3,119	1,111	2,806	0,356
TcCLB.506859.170	hypothetical protein, conserved	3	0	6,0	0,7	0,000	3,672	0,000	0	2	0	3,672	0,000	0,000	0,000
TcCLB.506913.10	hypothetical protein, conserved	3	3	49,3	26,7	3,714	8,884	3,545	1	3	3	5,171	-0,169	-30,556	-0,033
TcCLB.506925.90	hypothetical protein, conserved	3	2	22,0	15,7	-2,461	3,278	1,904	0	2	1	5,738	4,365	1,315	0,761
TcCLB.506949.50	cytochrome c, putative	3	1	23,0	4,0	-5,207	2,068	2,484	1	2	1	7,275	7,691	0,946	1,057
TcCLB.506979.10	hypothetical protein, conserved	3	3	11,7	7,0	1,359	4,440	1,601	0	2	1	3,081	0,242	12,747	0,078
TcCLB.506989.210	U-box domain protein, putative	3	2	29,3	9,7	-2,033	2,271	2,086	0	2	1	4,304	4,119	1,045	0,957
TcCLB.507003.70	hypothetical protein, conserved	3	1	6,7	1,7	-2,149	2,509	1,393	0	2	0	4,657	3,541	1,315	0,760
TcCLB.507007.40	hypothetical protein, conserved	3	3	38,3	27,7	1,898	4,251	4,003	1	2	3	2,353	2,106	1,118	0,895
TcCLB.507007.50	hypothetical protein, conserved	3	2	13,7	8,3	-2,625	6,274	1,624	0	3	0	8,899	4,249	2,094	0,478
TcCLB.507019.110	hypothetical protein	3	1	18,7	4,3	3,153	6,828	1,591	2	3	0	3,675	-1,561	-2,353	-0,425
TcCLB.507069.160	trans-sialidase, Group V, putative	3	3	33,7	10,3	1,673	3,983	2,587	1	2	2	2,310	0,914	2,527	0,396
TcCLB.507091.10	hypothetical protein, conserved	3	2	6,0	5,3	-1,973	2,185	2,146	0	2	2	4,158	4,120	1,009	0,991
	dihydroceramide synthase, putative														
TcCLB.507395.10	polypeptide deformylase-like protein, putative	3	0	7,0	2,0	0,000	3,086	0,000	1	2	0	3,086	0,000	0,000	0,000
	nucleotide-binding protein, putative														
TcCLB.507509.50	putative	3	1	6,7	3,3	0,000	3,236	6,963	0	3	1	3,236	6,963	0,465	2,152
TcCLB.507517.90	hypothetical protein, conserved	3	1	22,3	9,7	0,000	8,051	2,152	0	3	1	8,051	2,152	3,741	0,267
TcCLB.507519.90	RNA editing complex protein	3	2	11,0	6,3	-2,633	4,015	3,959	0	2	1	6,648	6,592	1,008	0,992
	MP63, putative														
TcCLB.507611.398	hypothetical protein, conserved	3	0	5,0	1,3	0,000	2,830	0,000	1	2	0	2,830	0,000	0,000	0,000
TcCLB.507625.190	protein phosphatase, putative	3	3	15,0	9,0	3,199	5,926	2,194	2	2	1	2,727	-1,005	-2,714	-0,368
TcCLB.507669.80	hypothetical protein, conserved	3	2	9,3	4,7	0,000	2,410	1,569	0	3	0	2,410	1,569	1,536	0,651
TcCLB.507673.30	protein phosphatase 1, regulatory subunit, putative	3	2	15,7	8,0	1,020	4,422	2,292	0	3	1	3,402	1,272	2,675	0,374
												10,29			
TcCLB.507867.60		3	2	13,7	9,0	-1,151	3,125	9,145	1	3	2	4,276	6	0,415	2,408

TcCLB.507947.10		3	3	18,3	15,0	2,064	4,303	2,196	2	2	2	2,239	0,132	17,021	0,059
TcCLB.507963.30	hypothetical protein, conserved	3	2	15,0	4,7	1,240	5,687	3,493	0	3	1	4,448	2,254	1,973	0,507
TcCLB.507979.50	hypothetical protein, conserved	3	1	13,0	3,7	-6,903	2,258	2,204	0	2	1	9,160	9,107	1,006	0,994
TcCLB.508045.40	hypothetical protein, conserved	3	0	6,0	2,3	0,000	2,881	0,000	0	2	0	2,881	0,000	0,000	0,000
TcCLB.508059.10	hypothetical protein, conserved	3	2	7,7	6,7	-2,472	2,910	2,965	1	3	1	5,381	5,436	0,990	1,010
TcCLB.508153.570	protein kinase, putative, serine/threonine kinase, putative	3	1	10,3	3,3	0,000	4,528	2,981	0	2	1	4,528	2,981	1,519	0,658
TcCLB.508153.670	hypothetical protein, conserved	3	2	13,0	7,0	1,342	3,892	2,506	0	3	1	2,550	1,164	2,191	0,456
TcCLB.508153.690	hypothetical protein	3	0	5,3	0,3	0,000	5,478	0,000	0	2	0	5,478	0,000	0,000	0,000
TcCLB.508175.290	hypothetical protein, conserved	3	1	14,0	10,0	-1,275	2,976	2,816	0	2	1	4,250	4,091	1,039	0,962
TcCLB.508207.110	60S ribosomal protein L7, putative	3	1	52,3	12,7	1,844	6,845	7,520	1	2	1	5,001	5,675	0,881	1,135
TcCLB.508207.31	hypothetical protein	3	2	15,0	11,7	-1,674	2,266	1,350	0	2	0	3,940	3,023	1,303	0,767
TcCLB.508211.50	hypothetical protein, conserved	3	3	21,7	13,7	-1,500	3,052	2,132	0	2	2	4,553	3,632	1,253	0,798
TcCLB.508231.230	protein transport protein Sec23A, putative	3	3	22,0	23,7	-1,192	2,012	2,030	0	2	2	3,204	3,222	0,994	1,006
TcCLB.508273.70	hypothetical protein, conserved	3	1	5,3	2,0	0,000	2,766	2,321	0	2	1	2,766	2,321	1,192	0,839
TcCLB.508307.125	hypothetical protein, conserved	3	2	10,3	6,7	1,305	3,571	1,458	0	2	0	2,266	0,153	14,838	0,067
TcCLB.508307.160	hypothetical protein, conserved	3	1	20,3	9,0	-2,160	2,447	1,620	0	2	0	4,607	3,780	1,219	0,820
TcCLB.508307.190	hypothetical protein, conserved	3	0	11,0	2,0	-2,179	2,798	0,000	0	3	0	4,977	2,179	2,284	0,438
TcCLB.508347.110	hypothetical protein, conserved	3	0	9,0	1,7	5,381	7,934	0,000	2	2	0	2,553	-5,381	-0,475	-2,107
TcCLB.508355.124	hypothetical protein, conserved	3	3	6,7	6,0	1,645	4,119	2,130	1	3	2	2,474	0,485	5,099	0,196
TcCLB.508409.170	hypothetical protein, conserved	3	1	8,7	2,3	0,000	3,778	2,484	0	3	1	3,778	2,484	1,521	0,657
TcCLB.508465.20	hypothetical protein, conserved	3	1	9,3	3,3	0,000	3,757	3,673	1	3	1	3,757	3,673	1,023	0,978
TcCLB.508567.80	DNA repair and recombination helicase protein PIF7, putative	3	3	21,7	13,3	-2,441	2,737	1,720	1	2	0	5,178	4,162	1,244	0,804
TcCLB.508645.4		3	2	5,0	1,7	0,000	2,179	1,595	0	2	1	2,179	1,595	1,366	0,732
TcCLB.508661.10	hypothetical protein, conserved	3	2	7,0	1,3	1,433	6,989	1,628	0	2	0	5,557	0,195	28,490	0,035
TcCLB.508691.10	hypothetical protein, conserved	3	1	5,7	3,3	1,207	5,112	1,863	0	2	0	3,905	0,656	5,953	0,168
TcCLB.508719.30	hypothetical protein, conserved	3	2	7,0	5,0	0,000	2,935	1,482	1	2	0	2,935	1,482	1,980	0,505
TcCLB.508777.110	hypothetical protein, conserved	3	0	15,3	6,3	-3,147	5,176	0,000	0	3	0	8,323	3,147	2,645	0,378
TcCLB.508777.150	hypothetical protein	3	2	34,7	15,3	-1,428	5,215	5,849	0	3	2	6,644	7,278	0,913	1,095
TcCLB.508787.30	pumilio/PUF RNA binding protein 3, putative (PUF3)	3	1	6,3	3,3	2,590	4,810	1,469	1	3	0	2,220	-1,121	-1,982	-0,505
TcCLB.508831.66	hypothetical protein, conserved	3	2	8,0	2,7	0,000	2,988	1,944	0	2	1	2,988	1,944	1,537	0,650
TcCLB.508901.10	hypothetical protein, conserved	3	3	23,0	10,7	-1,412	3,859	1,670	0	2	1	5,271	3,082	1,710	0,585
TcCLB.508919.100	hypothetical protein, conserved	3	2	6,3	4,0	-1,620	2,255	4,352	0	2	2	3,876	5,972	0,649	1,541
TcCLB.509049.51	hypothetical protein, conserved	3	3	18,0	8,7	1,613	8,193	4,918	1	2	3	6,581	3,305	1,991	0,502

TcCLB.509073.80	hypothetical protein, conserved	3	1	5,7	1,3	0,000	7,177	4,408	0	3	1	7,177	4,408	1,628	0,614
	kinetoplast														
	polyadenylation/uridylation factor														
TcCLB.509123.30	1	3	2	8,3	4,3	0,000	7,803	6,817	1	2	1	7,803	6,817	1,145	0,874
TcCLB.509137.10	hypothetical protein	3	2	7,3	2,7	0,000	3,957	6,966	0	3	2	3,957	6,966	0,568	1,760
TcCLB.509137.20	hypothetical protein	3	2	16,0	8,3	-2,257	2,277	3,619	0	2	1	4,534	5,876	0,772	1,296
	S-adenosylmethionine														
	decarboxylase														
	proenzyme,														
TcCLB.509167.110	putative	3	3	74,7	55,7	2,747	6,170	1,625	3	2	0	3,423	-1,121	-3,053	-0,328
												10,57			
TcCLB.509215.19	hypothetical protein, conserved	3	3	9,0	9,7	-6,050	4,523	1,942	1	2	2	2	7,991	1,323	0,756
												23,31			
TcCLB.509237.100	hypothetical protein, conserved	3	1	619,3	55,7	1,705	25,024	3,190	1	3	1	9	1,485	15,708	0,064
TcCLB.509253.40	hypothetical protein, conserved	3	2	28,3	10,3	-1,474	2,741	1,901	0	2	1	4,214	3,375	1,249	0,801
TcCLB.509331.20	hypothetical protein, conserved	3	3	9,7	10,0	-1,523	4,726	5,180	0	3	3	6,250	6,703	0,932	1,072
TcCLB.509429.100	hypothetical protein, conserved	3	1	9,7	4,3	-2,413	2,119	6,963	1	2	1	4,532	9,375	0,483	2,069
TcCLB.509429.230	hypothetical protein, conserved	3	3	5,0	2,0	-1,412	3,167	1,186	0	2	0	4,579	2,598	1,763	0,567
	retrotransposon hot spot (RHS)														
TcCLB.509463.41	protein, putative	3	0	12,3	5,7	0,000	2,615	0,000	0	2	0	2,615	0,000	0,000	0,000
TcCLB.509551.120	hypothetical protein	3	1	6,7	2,7	0,000	2,573	1,393	1	2	0	2,573	1,393	1,848	0,541
TcCLB.509571.40	hypothetical protein, conserved	3	1	9,0	5,7	1,214	4,774	1,218	0	2	0	3,560	0,005	770,608	0,001
TcCLB.509641.70	hypothetical protein, conserved	3	3	19,3	14,7	-2,412	3,527	1,851	0	3	1	5,939	4,262	1,393	0,718
TcCLB.509671.172	hypothetical protein, conserved	3	1	8,3	4,7	-1,658	2,607	1,790	1	2	0	4,265	3,448	1,237	0,809
												10,14			
TcCLB.509733.50	hypothetical protein, conserved	3	2	17,3	14,3	0,000	10,144	2,747	0	2	1	4	2,747	3,693	0,271
	rab-like GTPase activating protein,														
TcCLB.509789.30	putative	3	0	11,7	5,0	-1,672	2,368	0,000	0	3	0	4,040	1,672	2,416	0,414
TcCLB.509791.20	hypothetical protein, conserved	3	2	14,7	8,7	-2,045	2,359	3,081	0	2	1	4,404	5,126	0,859	1,164
TcCLB.509801.10	hypothetical protein, conserved	3	1	10,3	3,3	-1,346	2,633	1,065	0	2	0	3,980	2,411	1,651	0,606
TcCLB.509807.20	hypothetical protein, conserved	3	3	16,0	7,3	-1,310	2,505	8,269	0	2	3	3,815	9,579	0,398	2,511
TcCLB.509875.220	hypothetical protein, conserved	3	2	6,0	3,7	0,000	5,508	2,457	0	2	2	5,508	2,457	2,242	0,446
TcCLB.509911.110		3	2	6,0	9,0	0,000	2,244	1,378	0	2	0	2,244	1,378	1,629	0,614
												10,56			
TcCLB.509963.10	hypothetical protein, conserved	3	3	14,3	10,7	-6,922	3,643	2,086	0	2	1	6	9,009	1,173	0,853
TcCLB.509967.189	kinesin, putative	3	2	20,0	9,0	-1,555	2,413	2,077	0	2	1	3,968	3,632	1,093	0,915
TcCLB.509991.20	hypothetical protein, conserved	3	3	30,0	23,0	-1,476	3,197	1,482	1	2	0	4,673	2,958	1,580	0,633
TcCLB.510053.20	hypothetical protein, conserved	3	3	18,7	7,7	5,164	8,366	2,420	1	2	2	3,203	-2,744	-1,167	-0,857
	rev7, putative,mitotic spindle														
TcCLB.510091.40	checkpoint component, putative	3	2	9,7	2,7	0,000	2,157	1,247	0	2	0	2,157	1,247	1,729	0,578
TcCLB.510105.160		3	0	7,3	1,3	0,000	2,256	0,000	1	2	0	2,256	0,000	0,000	0,000
TcCLB.510127.70	leucine-rich repeat protein 1	3	3	9,3	5,3	2,236	4,629	2,297	2	3	2	2,394	0,062	38,825	0,026

(LRRP1), putative

TcCLB.510155.115	hypothetical protein, conserved	3	1	10,7	3,3	-1,729	2,466	2,204	0	2	1	4,195	3,933	1,066	0,938	
TcCLB.510181.140	hypothetical protein, conserved	3	3	9,7	5,0	2,766	5,883	5,771	2	3	3	3,117	3,005	1,037	0,964	
TcCLB.510187.234	tyrosine specific protein	3	1	8,0	3,0	-1,877	2,456	1,102	0	2	0	4,333	2,979	1,454	0,688	
	phosphatase, putative															
TcCLB.510187.270	COP-coated vesicle membrane protein erv25 precursor, putative,ER--golgi transport	3	3	17,7	10,3	1,793	4,193	1,810	1	2	1	2,401	0,018	136,180	0,007	
	protein erv25 precursor, putative															
TcCLB.510187.551	chaperonin HSP60, mitochondrial precursor, putative	3	1	24,3	9,3	-2,440	3,318	1,803	1	2	0	5,757	4,243	1,357	0,737	
TcCLB.510257.60	hypothetical protein, conserved	3	0	8,3	2,7	-2,029	8,128	0,000	1	3	0	10,157	2,029	5,005	0,200	
TcCLB.510309.30	PTP1-interacting protein, 39 kDa, putative	3	3	15,7	17,3	-1,442	4,051	2,043	0	3	2	5,493	3,485	1,576	0,634	
TcCLB.510329.250		3	0	6,7	2,0	0,000	2,145	0,000	0	2	0	2,145	0,000	0,000	0,000	
TcCLB.510349.20	hypothetical protein, conserved	3	1	78,0	16,3	1,267	4,828	4,775	0	2	1	3,560	3,508	1,015	0,985	
TcCLB.510351.80	hypothetical protein, conserved	3	0	8,0	5,0	-1,768	4,496	0,000	0	2	0	6,264	1,768	3,543	0,282	
TcCLB.510391.20	mucin-associated surface protein (MASP), putative	3	2	16,0	4,7	4,942	8,099	3,023	1	2	2	3,157	-1,919	-1,645	-0,608	
TcCLB.510403.89	synaptojanin (N-terminal domain), putative	3	2	12,0	5,7	-1,444	6,805	4,259	0	2	1	8,249	5,703	1,446	0,691	
TcCLB.510421.264	hypothetical protein, conserved	3	2	7,3	1,7	-1,702	4,079	1,441	0	2	0	5,781	3,143	1,839	0,544	
TcCLB.510507.50	RNA-binding protein, putative	3	2	8,0	5,3	2,514	6,556	1,733	2	2	1	4,042	-0,781	-5,178	-0,193	
TcCLB.510551.50	trans-sialidase, putative	3	2	53,7	4,3	0,000	6,268	1,783	0	2	1	6,268	1,783	3,516	0,284	
TcCLB.510599.40		3	2	6,3	4,3	0,000	2,244	2,013	1	2	1	2,244	2,013	1,114	0,897	
TcCLB.510609.110	ATP-dependent RNA helicase, putative	3	2	29,3	9,7	-4,002	2,512	1,843	0	2	1	6,514	5,844	1,115	0,897	
TcCLB.510657.150	hypothetical protein	3	2	14,0	7,3	1,858	6,607	1,987	1	2	1	4,750	0,129	36,759	0,027	
TcCLB.510681.41	surface protease GP63, putative	3	1	10,7	1,0	-1,805	2,014	7,452	0	2	1	3,819	9,256	0,413	2,424	
TcCLB.510693.270	trans-sialidase, Group VI, putative	3	2	5,0	3,7	0,000	3,795	1,976	0	3	1	3,795	1,976	1,921	0,521	
TcCLB.510693.320	mucin-associated surface protein (MASP), putative	3	2	37,7	14,7	0,000	21,933	3,488	1	2	2	21,933	3	3,488	6,289	0,159
TcCLB.510729.40	hypothetical protein, conserved	3	2	31,0	24,3	-3,325	3,152	1,861	0	2	1	6,477	5,187	1,249	0,801	
TcCLB.510763.70	hypothetical protein, conserved	3	1	5,7	3,7	0,000	2,347	1,114	1	2	0	2,347	1,114	2,107	0,475	
TcCLB.510863.20	hypothetical protein	3	1	7,7	4,0	-2,781	3,104	3,249	1	2	1	5,885	6,030	0,976	1,025	
TcCLB.510903.30	helicase-like protein, putative	3	3	23,3	12,7	-1,047	2,139	1,847	0	2	1	3,186	2,894	1,101	0,908	
TcCLB.510903.70	hypothetical protein, conserved	3	2	14,3	5,3	-3,348	5,901	2,239	0	2	1	9,249	5,587	1,655	0,604	
TcCLB.510965.5	elongation factor 2, putative	3	2	10,0	3,3	0,000	3,216	1,758	1	3	0	3,216	1,758	1,829	0,547	
TcCLB.510993.10	(H )-ATPase G subunit	3	3	25,0	16,7	-1,632	2,401	1,801	1	2	1	4,033	3,433	1,175	0,851	
TcCLB.510999.10	hypothetical protein, conserved	3	0	52,3	9,0	-2,176	2,850	0,000	0	2	0	5,026	2,176	2,309	0,433	

TcCLB.511003.90	hypothetical protein, conserved	3	0	8,7	4,3	-2,498	2,321	0,000	1	2	0	4,819	2,498	1,929	0,518
TcCLB.511031.10	hypothetical protein, conserved	3	3	13,3	6,3	1,207	3,653	2,829	0	2	2	2,446	1,622	1,508	0,663
TcCLB.511153.60	palmitoyl acyltransferase 9, putative	3	3	6,3	5,3	-2,269	4,317	1,866	0	2	1	6,586	4,135	1,593	0,628
TcCLB.511167.60	hypothetical protein, conserved	3	2	34,0	20,7	-1,528	4,405	2,629	0	3	1	5,933	4,157	1,427	0,701
TcCLB.511169.29	tubulin-tyrosine ligase-like protein, putative	3	1	21,7	6,0	0,000	2,168	1,218	1	2	0	2,168	1,218	1,779	0,562
TcCLB.511181.70	hypothetical protein, conserved	3	1	21,0	0,7	0,000	9,428	1,393	1	2	0	9,428	1,393	6,770	0,148
TcCLB.511211.230	hypothetical protein, conserved	3	0	6,7	5,0	-1,976	3,944	0,000	0	3	0	5,920	1,976	2,996	0,334
TcCLB.511215.10	hypothetical protein, conserved	3	2	33,0	15,7	-1,217	3,313	1,889	0	2	1	4,530	3,106	1,459	0,686
TcCLB.511245.10	hypothetical protein, conserved	3	2	59,7	21,0	-1,613	6,119	6,136	0	2	2	7,732	7,749	0,998	1,002
TcCLB.511285.10	hypothetical protein	3	2	23,0	13,3	1,348	3,701	1,353	0	3	0	2,354	0,005	438,924	0,002
TcCLB.511287.30	60S ribosomal protein L17, putative	3	1	31,0	6,0	-7,897	2,331	2,387	0	2	1	10,228	10,285	0,995	1,006
TcCLB.511287.90	ribosome biogenesis protein, putative	3	2	33,7	19,7	-1,274	3,130	2,621	0	2	2	4,404	3,895	1,131	0,884
TcCLB.511303.140	hypothetical protein, conserved	3	1	5,7	1,7	0,000	7,002	3,481	1	3	1	7,002	3,481	2,011	0,497
TcCLB.511303.50	hypothetical protein, conserved	3	2	9,3	3,0	2,934	9,222	3,191	2	3	2	6,288	0,257	24,478	0,041
TcCLB.511355.30	phosphatidic acid phosphatase protein, putative	3	2	10,7	8,7	2,622	9,200	3,195	2	2	1	6,579	0,573	11,475	0,087
TcCLB.511407.50	hypothetical protein, conserved	3	1	16,0	8,7	-3,830	4,081	1,163	0	2	0	7,911	4,993	1,584	0,631
TcCLB.511445.30	hypothetical protein, conserved	3	2	12,3	6,0	-2,009	3,488	2,154	0	2	1	5,497	4,163	1,320	0,757
TcCLB.511469.10	hypothetical protein, conserved	3	0	6,7	2,3	-3,017	3,272	0,000	0	2	0	6,289	3,017	2,084	0,480
TcCLB.511507.24	hypothetical protein, conserved	3	3	10,3	10,3	-2,271	2,984	1,874	0	2	2	5,256	4,146	1,268	0,789
TcCLB.511509.50	hypothetical protein, conserved	3	2	15,0	11,3	4,000	7,022	2,120	2	2	1	3,022	-1,880	-1,607	-0,622
TcCLB.511541.40	hypothetical protein, conserved	3	3	21,3	18,3	4,074	7,239	3,356	2	3	2	3,165	-0,719	-4,405	-0,227
TcCLB.511555.20	hypothetical protein, conserved	3	2	10,0	2,3	1,651	3,953	2,224	1	2	1	2,303	0,573	4,018	0,249
TcCLB.511585.60	hypothetical protein, conserved	3	2	141,3	17,7	1,568	4,942	4,913	0	2	1	3,375	3,346	1,009	0,991
TcCLB.511627.120	aldo/keto reductase, putative	3	1	22,7	11,7	-1,469	2,026	2,112	0	2	1	3,494	3,581	0,976	1,025
TcCLB.511717.80	DNA repair and recombination helicase protein PIF4, putative	3	2	16,3	17,3	1,732	5,956	1,363	0	3	0	4,224	-0,369	-11,436	-0,087
TcCLB.511727.129	ribosomal protein L18, putative	3	3	44,3	43,7	1,085	3,211	2,324	0	2	2	2,125	1,239	1,715	0,583
TcCLB.511737.50	hypothetical protein, conserved	3	2	26,3	7,3	1,499	4,094	4,123	0	3	2	2,595	2,624	0,989	1,011
TcCLB.511751.140	hypothetical protein	3	3	51,7	28,3	2,717	7,377	4,215	1	3	2	4,660	1,499	3,110	0,322
TcCLB.511751.160	hypothetical protein, conserved	3	1	10,7	3,3	2,470	6,221	1,242	2	2	0	3,751	-1,228	-3,055	-0,327
TcCLB.511809.20	methylenetetrahydrofolate dehydrogenase-like protein, putative	3	2	15,3	10,7	-3,316	2,482	2,195	1	2	1	5,798	5,511	1,052	0,950
TcCLB.511817.310	hypothetical protein, conserved	3	2	17,7	10,0	-1,608	2,067	1,432	0	2	0	3,675	3,040	1,209	0,827

TcCLB.511827.60	CCR4 associated factor, putative	3	0	8,7	3,0	-2,479	2,124	0,000	0	2	0	4,604	2,479	1,857	0,539
								18,28				13,95	14,17		
TcCLB.511855.20	ubiquitin-like protein, putative	3	2	322,3	340,3	4,108	18,062	3	2	3	2	3	5	0,984	1,016
TcCLB.511891.10	trans-sialidase, Group II, putative	3	3	9,0	7,0	2,038	4,997	7,027	1	3	3	2,960	4,989	0,593	1,686

Identificação alvos TcRBP40	Anotação dos alvos de TcRBP40	Filtro 1: Número de réplicas enriquecidas na fração ligada em TcRBP7	Filtro 1: Número de réplicas enriquecidas na fração ligada em TcRBP40	Filtro 2: Média de leituras da fração ligada em TcRBP7	Filtro 2: Média de leituras da fração ligada em TcRBP40	Filtro 3: Média FC em Controle	Filtro 3: Média FC em TcRBP7	Filtro 3: Média FC em TcRBP40	Filtro 4: Número de réplicas com FC ≥ 2 em Controle	Filtro 4: Número de réplicas com FC ≥ 2 em TcRBP7	Filtro 4: Número de réplicas com FC ≥ 2 em TcRBP40	Filtro 5: Δ FC Médio (FC TcRBP7 - FC Controle)	Filtro 5: Δ FC Médio (FC TcRBP40 - FC Controle)	FC RBP7 / FC RBP40	FC RBP40 / FC RBP7
TcCLB.419043.10	hypothetical protein, conserved	2	3	48,7	38,3	4,319	3,256	6,640	3	2	3	-1,063	2,321	-0,458	-2,184
TcCLB.420369.40	hypothetical protein, conserved	1	3	7,0	6,7	-2,684	1,017	4,227	0	0	2	3,701	6,912	0,535	1,867
TcCLB.424123.20	hypothetical protein, conserved	3	3	51,0	62,7	2,198	2,052	4,896	1	2	3	-0,145	2,699	-0,054	18,550
TcCLB.432027.9	hypothetical protein conserved	2	3	6,3	7,7	0,000	2,221	4,806	1	1	3	2,221	4,806	0,462	2,163
TcCLB.440101.10	fatty acid desaturase, putative	2	3	4,7	7,0	3,836	3,772	6,872	2	1	2	-0,064	3,036	-0,021	47,492
TcCLB.450161.10	cation transporter, putative	2	3	6,3	5,3	-2,400	1,532	3,283	0	0	3	3,932	5,684	0,692	1,445
TcCLB.452039.9	hypothetical protein, conserved	3	3	20,7	17,0	-1,566	2,055	3,729	0	1	2	3,621	5,295	0,684	1,462
TcCLB.464807.20	hypothetical protein, conserved	1	3	5,3	8,0	0,000	8,447	4,557	0	1	3	8,447	4,557	1,854	0,539
TcCLB.471901.10	hypothetical protein conserved	2	3	13,7	14,3	1,311	1,868	3,394	0	1	2	0,557	2,083	0,267	3,740
TcCLB.476733.10	hypothetical protein, conserved	1	3	10,7	7,7	-2,945	3,054	5,838	0	1	2	5,999	8,783	0,683	1,464
TcCLB.480975.10	hypothetical protein, conserved	3	3	8,0	6,0	-3,462	2,988	6,017	0	2	3	6,450	9,480	0,680	1,470
TcCLB.503399.10	hypothetical protein conserved	0	3	3,0	5,0	0,000	0,000	3,014	0	0	3	0,000	3,014	0,000	0,000
TcCLB.503419.20	adenosine deaminase-like protein, putative (ADAT1h)	1	3	2,3	5,7	-1,668	1,212	2,740	1	0	2	2,880	4,408	0,653	1,531
TcCLB.503461.20	per-hexamer repeat protein 5, putative	1	3	4,7	10,7	-2,537	1,408	3,515	0	0	2	3,945	6,052	0,652	1,534
TcCLB.503487.70	hypothetical protein conserved	2	3	30,7	22,3	1,305	3,549	4,029	0	1	2	2,244	2,724	0,824	1,214
TcCLB.503559.109	electron-transfer-flavoprotein, alpha polypeptide, putative	2	3	7,3	6,0	0,000	2,893	3,425	1	1	2	2,893	3,425	0,845	1,184
TcCLB.503613.20	proteasome alpha 7 subunit, putative	3	3	12,3	15,7	0,000	3,068	4,053	0	2	3	3,068	4,053	0,757	1,321
TcCLB.503635.20	hypothetical protein, conserved	2	3	6,3	9,3	-1,444	1,935	2,975	1	1	2	3,379	4,419	0,765	1,308
TcCLB.503641.19	peroxin 12, putative (PEX12)	0	3	2,3	5,0	0,000	0,000	3,750	0	0	2	0,000	3,750	0,000	0,000

TcCLB.503657.50	hypothetical protein, conserved	1	3	13,3	14,3	-1,812	4,399	2,440	0	1	2	6,212	4,253	1,461	0,685
TcCLB.503657.9	hypothetical protein, conserved	3	3	49,0	31,3	-1,142	2,883	2,587	0	2	2	4,024	3,729	1,079	0,927
TcCLB.503685.20	hypothetical protein, conserved	1	3	11,0	15,0	-1,430	2,112	3,044	0	1	2	3,541	4,474	0,792	1,263
TcCLB.503723.70	protein farnesyltransferase, putative	1	3	3,3	6,3	0,000	1,408	7,886	0	0	3	1,408	7,886	0,179	5,601
TcCLB.503779.20	hypothetical protein conserved	1	3	16,0	9,7	2,291	2,816	5,326	2	1	2	0,524	3,035	0,173	5,790
TcCLB.503839.19	hypothetical protein, conserved	1	3	6,0	7,3	1,987	2,828	4,312	1	1	2	0,841	2,325	0,362	2,765
TcCLB.503841.10	proteasome activator protein pa26, putative	2	3	8,3	14,0	3,829	1,560	5,874	1	0	3	-2,269	2,045	-1,109	-0,901
TcCLB.503865.70	hypothetical protein, conserved	2	3	29,7	23,3	-3,154	2,091	2,704	1	1	2	5,245	5,858	0,895	1,117
TcCLB.503867.5	hypothetical protein, conserved	1	3	3,3	5,0	0,000	3,817	2,027	0	1	2	3,817	2,027	1,883	0,531
TcCLB.503891.100	proteasome beta 2 subunit, putative	2	3	9,0	13,7	-1,954	1,182	5,331	0	0	2	3,135	7,285	0,430	2,323
TcCLB.503893.120	calcium uniporter protein, mitochondrial, putative	3	3	14,7	11,3	2,877	1,658	5,956	2	1	3	-1,219	3,079	-0,396	-2,525
TcCLB.503897.140	hypothetical protein, conserved	2	3	8,3	6,3	0,000	3,107	4,356	1	1	2	3,107	4,356	0,713	1,402
TcCLB.503939.70	hypothetical protein	1	3	24,7	33,7	-1,372	2,429	2,201	0	1	2	3,801	3,572	1,064	0,940
TcCLB.503965.10	hypothetical protein conserved	1	3	3,0	11,7	2,867	1,056	5,109	2	0	3	-1,811	2,242	-0,808	-1,238
TcCLB.503995.30	hypothetical protein	2	3	15,3	13,0	-2,473	5,746	7,603	0	2	3	8,220	10,076	0,816	1,226
TcCLB.504033.110	hypothetical protein, conserved	2	3	8,7	9,7	0,000	2,290	5,439	1	2	3	2,290	5,439	0,421	2,376
TcCLB.504041.20	hypothetical protein conserved	2	3	13,0	16,0	1,305	1,402	3,369	0	0	2	0,097	2,065	0,047	21,334
TcCLB.504057.110	ribosomal RNA processing protein 6, putative,exosome complex exonuclease, putative (RRP6)	3	3	187,3	95,0	2,140	5,310	5,009	2	2	3	3,170	2,869	1,105	0,905
TcCLB.504109.10	hypothetical protein, conserved	3	3	117,0	64,0	-2,007	3,625	2,571	0	3	2	5,632	4,578	1,230	0,813
TcCLB.504421.19	hypothetical protein conserved	3	3	7,7	6,3	1,468	2,666	7,690	0	2	3	1,198	6,222	0,193	5,192
TcCLB.504425.110	permease-like protein, putative	2	3	3,7	8,0	-2,014	4,474	4,062	0	1	3	6,488	6,076	1,068	0,936
TcCLB.504427.110	acetyl-CoA synthetase, putative	1	3	41,0	66,3	1,429	1,889	3,645	0	0	3	0,461	2,216	0,208	4,810
TcCLB.504741.40	hypothetical protein conserved	2	3	4,3	5,0	1,818	1,512	7,929	1	0	2	-0,306	6,111	-0,050	19,951
TcCLB.504867.10	hypothetical protein, conserved	1	3	11,0	8,7	0,000	3,681	2,126	0	1	2	3,681	2,126	1,731	0,578
TcCLB.504867.120	hypothetical protein	1	3	6,0	7,3	-2,855	1,527	2,852	0	0	3	4,382	5,707	0,768	1,303
TcCLB.505025.80	hypothetical protein, conserved	2	3	27,0	59,3	1,239	6,276	4,200	0	2	3	5,037	2,962	1,701	0,588
TcCLB.505919.40	protein kinase, putative	3	3	45,0	41,7	-3,284	3,021	2,151	1	2	2	6,305	5,435	1,160	0,862
TcCLB.506007.30	hypothetical protein, conserved	2	3	17,7	31,0	1,994	6,709	10,578	1	2	3	4,715	8,584	0,549	1,821
TcCLB.506165.4	hypothetical protein, conserved	1	3	4,7	11,3	0,000	7,743	5,153	1	1	3	7,743	5,153	1,503	0,666
TcCLB.506181.50	60S ribosomal protein L18, putative	1	3	3,3	6,0	0,000	1,909	4,336	0	0	3	1,909	4,336	0,440	2,272
TcCLB.506203.100	hypothetical protein	2	3	8,3	9,0	-1,449	1,778	4,407	1	1	2	3,227	5,857	0,551	1,815
TcCLB.506213.40	hypothetical protein conserved	2	3	6,0	12,7	0,000	1,705	3,275	0	1	2	1,705	3,275	0,521	1,921
TcCLB.506221.110	chromosomal passenger protein, putative (CPC2)	3	3	12,3	9,3	5,279	1,888	8,688	2	1	3	-3,391	3,410	-0,994	-1,006
TcCLB.506223.20	hypothetical protein conserved	2	3	10,7	49,3	3,941	1,824	22,925	2	1	3	-2,117	18,984	-0,112	-8,966



TcCLB.506251.120	hypothetical protein conserved	3	3	9,0	14,3	2,942	1,693	8,895	3	0	3	-1,249	5,953	-0,210	-4,767
TcCLB.506257.10	DNA-RNA non-specific endonuclease protein-like putative	2	3	7,7	9,0	-2,297	1,528	3,169	1	0	2	3,825	5,466	0,700	1,429
TcCLB.506265.90	hypothetical protein, conserved mucin-associated surface protein (MASP), putative	1	3	6,3	8,0	-2,755	2,251	3,741	0	1	2	5,005	6,496	0,771	1,298
TcCLB.506335.100	hypothetical protein conserved	3	3	8,0	6,7	0,000	3,278	2,639	0	2	2	3,278	2,639	1,242	0,805
TcCLB.506351.30	soluble N-ethylmaleimide sensitive factor (NSF) attachment protein, putative	0	3	2,7	5,0	-1,399	0,000	2,052	0	0	2	1,399	3,451	0,405	2,466
TcCLB.506363.150	putative	1	3	5,7	8,3	2,934	2,624	7,240	2	1	3	-0,309	4,306	-0,072	-
TcCLB.506401.240	hypothetical protein, conserved	3	3	21,7	38,3	-1,232	1,670	2,625	0	1	3	2,902	3,857	0,752	1,329
TcCLB.506409.230	hypothetical protein, conserved mitochondrial pyruvate carrier	3	3	42,7	31,7	9,470	22,674	15,640	2	3	3	13,204	6,170	2,140	0,467
TcCLB.506443.40	protein 2, putative	2	3	26,0	5,7	-1,444	7,871	2,549	1	2	2	9,315	3,992	2,333	0,429
TcCLB.506443.60	hypothetical protein	3	3	28,3	15,0	-1,519	1,730	3,616	0	1	2	3,248	5,135	0,633	1,581
TcCLB.506469.10	hypothetical protein, conserved	3	3	14,3	14,3	0,000	3,873	12,203	0	1	3	3,873	12,203	0,317	3,151
TcCLB.506469.130	SNARE domain-containing protein, putative	1	3	7,7	10,7	4,115	3,754	7,751	2	1	3	-0,360	3,636	-0,099	-
TcCLB.506479.110	putative	3	3	18,3	15,0	-2,525	1,785	2,165	0	1	2	4,310	4,690	0,919	1,088
TcCLB.506483.10	hypothetical protein	3	3	74,3	41,7	2,625	1,817	8,264	1	1	3	-0,808	5,639	-0,143	-6,980
TcCLB.506503.100	class 3 lipase, putative	2	3	6,0	8,0	-1,310	1,202	2,100	1	0	2	2,512	3,410	0,737	1,357
TcCLB.506513.190	U3 snoRNA-associated protein UTP11, putative	1	3	11,3	23,7	0,000	4,047	2,464	0	1	2	4,047	2,464	1,642	0,609
TcCLB.506525.10	NADH-ubiquinone oxidoreductase, mitochondrial, putative	2	3	11,0	16,0	-3,292	1,310	2,228	0	0	2	4,602	5,520	0,834	1,199
TcCLB.506525.50	hypothetical protein, conserved	3	3	7,3	11,0	-1,588	1,952	2,331	1	2	2	3,540	3,919	0,903	1,107
TcCLB.506525.80	hypothetical protein, conserved	2	3	12,3	14,7	1,188	2,060	3,268	0	1	2	0,872	2,080	0,419	2,386
TcCLB.506637.90	N-myristoyl transferase, putative	1	3	9,0	12,7	-2,365	3,499	3,340	1	1	3	5,864	5,704	1,028	0,973
TcCLB.506657.88	hypothetical protein, conserved	1	3	13,7	10,3	1,311	2,514	3,348	0	1	2	1,203	2,037	0,590	1,694
TcCLB.506661.160	hypothetical protein conserved	3	3	70,7	24,3	-2,239	16,463	6,260	0	2	2	18,702	8,499	2,201	0,454
TcCLB.506693.10	hypothetical protein, conserved	3	3	10,7	23,3	0,000	6,641	7,431	0	3	2	6,641	7,431	0,894	1,119
TcCLB.506735.30	hypothetical protein	2	3	4,3	6,7	0,000	1,193	3,540	1	0	3	1,193	3,540	0,337	2,969
TcCLB.506743.10	hypothetical protein conserved	2	3	7,0	6,0	0,000	1,700	2,451	0	1	2	1,700	2,451	0,693	1,442
TcCLB.506755.110	hypothetical protein	2	3	12,0	13,3	1,956	1,895	5,302	1	1	2	-0,060	3,346	-0,018	-
TcCLB.506773.170	hypothetical protein conserved	3	3	10,7	8,3	-1,497	7,150	3,111	1	1	2	8,647	4,608	1,876	0,533
TcCLB.506775.130	hypothetical protein, conserved	1	3	7,0	16,7	-6,146	1,088	2,951	0	0	2	7,234	9,097	0,795	1,258
TcCLB.506811.160	hypothetical protein, conserved	1	3	9,3	12,7	-3,080	1,119	2,328	0	0	2	4,199	5,408	0,776	1,288
TcCLB.506813.5	R27-2 protein, putative	2	3	14,0	22,7	-1,249	1,570	2,491	0	0	3	2,819	3,740	0,754	1,327
TcCLB.506821.80	hypothetical protein	3	3	46,0	87,7	3,182	2,394	5,887	3	1	3	-0,788	2,705	-0,291	-3,432
TcCLB.506825.80	hypothetical protein conserved	0	3	2,7	5,0	0,000	0,000	4,055	0	0	2	0,000	4,055	0,000	0,000

TcCLB.506831.20	amino acid permease, putative	2	3	5,7	9,3	3,083	2,044	9,514	2	1	3	-1,039	6,430	-0,162	-6,187
TcCLB.506855.190	hypothetical protein, conserved	2	3	25,7	30,0	-5,515	1,169	2,490	1	0	2	6,683	8,005	0,835	1,198
TcCLB.506855.290	hypothetical protein, conserved proteasome regulatory ATPase subunit 1, putative	1	3	11,7	12,7	1,776	2,077	4,364	1	1	2	0,301	2,587	0,116	8,593
TcCLB.506857.90	hypothetical protein	1	3	48,3	74,0	-1,569	1,222	3,354	1	0	2	2,791	4,923	0,567	1,764
TcCLB.506859.120	hypothetical protein, conserved	1	3	5,0	10,7	4,642	1,215	9,719	1	0	3	-3,427	5,077	-0,675	-1,481
TcCLB.506885.330	hypothetical protein, conserved	1	3	5,3	8,7	-2,815	3,817	3,148	0	1	2	6,633	5,963	1,112	0,899
TcCLB.506891.10	hypothetical protein, conserved	3	3	3,7	5,0	-1,609	1,906	4,362	1	1	2	3,515	5,971	0,589	1,699
TcCLB.506907.14	hypothetical protein, conserved	2	3	7,0	11,0	-3,099	6,120	3,949	0	2	3	9,218	7,048	1,308	0,765
TcCLB.506943.20	hypothetical protein, conserved	2	3	9,7	7,7	-2,354	1,146	2,540	0	0	2	3,500	4,894	0,715	1,398
TcCLB.506957.130	hypothetical protein, conserved 60S ribosomal protein L18a, putative	2	3	18,0	16,0	-1,225	2,211	2,907	0	2	2	3,436	4,133	0,832	1,203
TcCLB.506957.90	hypothetical protein, conserved	2	3	41,0	41,0	1,301	4,540	3,566	0	2	3	3,239	2,265	1,430	0,699
TcCLB.507007.40	hypothetical protein, conserved	3	3	38,3	27,7	1,898	4,251	4,003	1	2	3	2,353	2,106	1,118	0,895
TcCLB.507011.90	hypothetical protein conserved	2	3	3,7	12,0	1,919	2,037	9,873	1	1	2	0,119	7,955	0,015	67,004
TcCLB.507017.159	hypothetical protein, conserved	1	3	3,0	5,3	-1,156	2,828	6,176	0	1	3	3,984	7,332	0,543	1,841
TcCLB.507047.90	hypothetical protein mitochondrial RNA binding protein, putative	2	3	7,7	8,3	-1,949	1,822	2,838	1	1	3	3,771	4,787	0,788	1,269
TcCLB.507049.100	mucin-associated surface protein (MASP), putative	0	3	4,0	8,0	0,000	0,000	5,474	1	0	2	0,000	5,474	0,000	0,000
TcCLB.507071.180	hypothetical protein	0	3	2,7	10,0	1,752	0,000	4,130	1	0	3	-1,752	2,378	-0,737	-1,357
TcCLB.507071.340	hypothetical protein aminopeptidase P1, putative, metallo-peptidase, Clan MG, Family M24, putative	3	3	45,0	52,0	1,184	1,837	3,590	0	1	3	0,654	2,406	0,272	3,682
TcCLB.507081.110	hypothetical protein conserved	1	3	14,7	20,7	-1,331	8,589	2,504	1	1	2	9,920	3,835	2,587	0,387
TcCLB.507099.120	hypothetical protein conserved beta galactofuranosyl glycosyltransferase, putative	1	3	7,7	6,0	2,722	4,927	5,328	2	1	3	2,205	2,605	0,846	1,182
TcCLB.507121.40	fumarate hydratase, putative	2	3	184,0	227,7	-1,790	1,608	2,392	0	0	2	3,399	4,182	0,813	1,231
TcCLB.507257.60	hypothetical protein, conserved	1	3	8,0	27,7	2,470	1,363	4,801	1	0	3	-1,107	2,331	-0,475	-2,106
TcCLB.507317.34	hypothetical protein conserved	1	3	6,0	8,0	0,000	5,162	13,599	1	1	2	5,162	13,599	0,380	2,634
TcCLB.507485.50	hypothetical protein, conserved	0	3	2,7	8,0	0,000	0,000	4,793	0	0	3	0,000	4,793	0,000	0,000
TcCLB.507515.110	hypothetical protein, conserved	1	3	4,7	5,3	-1,833	1,909	2,299	0	0	2	3,742	4,132	0,906	1,104
TcCLB.507515.120	sphingosine kinase A, B, putative	3	3	14,7	15,7	2,787	2,121	5,155	2	2	2	-0,666	2,368	-0,281	-3,554
TcCLB.507677.120	hypothetical protein conserved	0	3	1,7	5,0	0,000	0,000	2,052	1	0	2	0,000	2,052	0,000	0,000
TcCLB.507705.10	hypothetical protein, conserved	2	3	48,7	65,0	-1,177	3,382	2,235	1	1	2	4,558	3,412	1,336	0,748
TcCLB.507711.160	hypothetical protein, conserved	2	3	17,3	11,3	0,000	3,117	4,597	0	1	2	3,117	4,597	0,678	1,475
TcCLB.507711.40	RNA-binding protein, putative	3	3	5,0	7,7	2,871	4,109	5,367	2	3	3	1,238	2,496	0,496	2,017
TcCLB.507771.50	glycerolphosphate mutase, putative	3	3	16,7	14,3	-2,104	1,994	3,473	0	1	3	4,098	5,577	0,735	1,361
TcCLB.507773.50	hypothetical protein, conserved	2	3	15,3	12,0	-1,250	5,196	2,589	0	2	2	6,446	3,839	1,679	0,596
TcCLB.507803.30	hypothetical protein	2	3	9,3	12,7	-1,236	2,975	3,819	1	2	3	4,212	5,055	0,833	1,200

TcCLB.507811.114	hypothetical protein, conserved	2	3	3,3	5,0	0,000	5,201	2,617	0	2	2	5,201	2,617	1,987	0,503
TcCLB.507823.39	hypothetical protein conserved	2	3	6,7	9,0	0,000	1,913	6,733	1	1	2	1,913	6,733	0,284	3,520
TcCLB.507829.10	calpain-like cysteine peptidase, putative, cysteine peptidase, Clan CA, family C2, putative	3	3	8,3	9,3	2,419	4,373	8,440	1	2	2	1,954	6,021	0,324	3,082
TcCLB.507837.10	separase, putative, cysteine peptidase, Clan CD, family C50, putative	1	3	5,3	7,3	3,775	2,346	7,082	2	1	3	-1,428	3,308	-0,432	-2,316
TcCLB.507873.50	hypothetical protein, conserved	2	3	6,0	8,0	2,725	5,415	5,657	1	2	3	2,690	2,932	0,918	1,090
TcCLB.507883.100	malate dehydrogenase, putative	2	3	5,7	8,3	-6,838	1,381	7,317	0	0	2	8,219	14,155	0,581	1,722
TcCLB.507913.30	proliferative cell nuclear antigen (PCNA), putative	2	3	17,3	19,7	-2,541	2,133	2,424	0	2	2	4,674	4,965	0,941	1,062
TcCLB.507957.179	mucin-associated surface protein (MASP), putative	1	3	0,7	8,7	0,000	1,408	4,007	0	0	3	1,408	4,007	0,351	2,846
TcCLB.507993.270	hypothetical protein conserved	2	3	4,7	12,3	2,603	1,521	4,658	2	0	3	-1,082	2,055	-0,527	-1,898
TcCLB.508153.260	hypothetical protein, conserved	1	3	11,0	15,3	-1,380	1,207	2,278	0	0	3	2,586	3,657	0,707	1,414
TcCLB.508169.100	hypothetical protein conserved	2	3	6,0	10,3	2,394	2,965	13,066	2	2	2	0,570	10,671	0,053	18,713
TcCLB.508211.50	hypothetical protein, conserved	3	3	21,7	13,7	-1,500	3,052	2,132	0	2	2	4,553	3,632	1,253	0,798
TcCLB.508221.1010	mucin-associated surface protein (MASP), putative	2	3	5,0	10,0	4,150	2,992	6,975	2	1	3	-1,158	2,825	-0,410	-2,439
TcCLB.508231.230	protein transport protein Sec23A, putative	3	3	22,0	23,7	-1,192	2,012	2,030	0	2	2	3,204	3,222	0,994	1,006
TcCLB.508257.190	hypothetical protein, conserved	3	3	5,0	5,0	-1,347	2,415	2,938	1	1	2	3,763	4,285	0,878	1,139
TcCLB.508265.4	synaptojanin (N-terminal domain), putative	1	3	3,3	5,0	0,000	9,855	3,589	0	1	2	9,855	3,589	2,745	0,364
TcCLB.508269.30	translation initiation factor 2 subunit, putative	1	3	4,3	11,3	0,000	3,030	3,091	0	1	2	3,030	3,091	0,980	1,020
TcCLB.508347.100	hypothetical protein conserved	2	3	11,3	18,3	1,501	1,110	4,172	0	0	3	-0,391	2,672	-0,146	-6,837
TcCLB.508347.40	hypothetical protein conserved	0	3	9,3	11,3	-1,512	0,000	2,345	1	0	2	1,512	3,857	0,392	2,550
TcCLB.508355.104	hypothetical protein, conserved	2	3	10,3	5,0	0,000	3,106	6,792	1	1	3	3,106	6,792	0,457	2,187
TcCLB.508375.30	aspartate carbamoyltransferase, putative	1	3	2,3	5,3	0,000	1,877	4,261	1	0	3	1,877	4,261	0,441	2,270
TcCLB.508387.30	hypothetical protein conserved	1	3	6,0	6,0	0,000	1,564	3,449	0	0	2	1,564	3,449	0,454	2,205
TcCLB.508409.338	DNA repair protein, putative	1	3	2,7	5,0	0,000	1,760	3,651	1	0	3	1,760	3,651	0,482	2,075
TcCLB.508441.70	enoyl-CoA hydratase/Enoyl-CoA isomerase/3-hydroxyacyl-CoA dehydrogenase, putative	2	3	36,0	32,7	-1,093	2,110	2,182	0	1	2	3,203	3,276	0,978	1,023
TcCLB.508461.470	hypothetical protein, conserved	3	3	7,3	5,7	-1,530	1,929	2,067	0	2	2	3,459	3,597	0,962	1,040
TcCLB.508461.60	synaptojanin (N-terminal domain), putative, inositol/phosphatidylinositol phosphatase, putative	3	3	11,7	17,0	2,750	2,432	6,880	2	1	3	-0,317	4,131	-0,077	13,027
TcCLB.508507.59	NADH-ubiquinone oxidoreductase 20 kDa subunit, mitochondrial precursor, putative	1	3	7,3	9,7	1,738	1,091	5,966	0	0	2	-0,647	4,229	-0,153	-6,537
TcCLB.508541.130	mucin-associated surface protein (MASP), putative	1	3	4,7	18,7	0,000	3,232	4,675	0	1	3	3,232	4,675	0,691	1,447

TcCLB.508543.110	hypothetical protein, conserved	2	3	6,3	6,3	-1,267	4,599	3,398	1	1	2	5,866	4,665	1,257	0,795
TcCLB.508593.120	hypothetical protein conserved	3	3	11,7	18,7	1,941	1,117	4,253	1	0	2	-0,824	2,312	-0,356	-2,806
TcCLB.508707.227	ATP binding protein-like protein, putative	2	3	23,3	25,3	1,235	1,776	4,932	0	1	2	0,541	3,698	0,146	6,829
TcCLB.508723.24	hypothetical protein, conserved	1	3	5,3	5,7	0,000	6,059	3,133	1	1	2	6,059	3,133	1,934	0,517
TcCLB.508735.30	mucin-associated surface protein (MASP), putative	3	3	15,0	12,3	1,979	2,822	4,415	1	2	3	0,843	2,437	0,346	2,892
TcCLB.508817.150	hypothetical protein conserved	3	3	26,0	37,3	4,213	3,577	8,475	1	2	2	-0,635	4,262	-0,149	-6,708
TcCLB.508827.139	coatomer gamma subunit, putative	1	3	4,7	9,3	-1,611	4,241	2,888	1	1	2	5,852	4,498	1,301	0,769
TcCLB.508827.20	hypothetical protein, conserved	2	3	10,7	15,0	-2,561	2,253	4,375	0	2	3	4,814	6,936	0,694	1,441
TcCLB.508879.200	hypothetical protein conserved	2	3	7,3	7,0	0,000	1,993	4,490	1	1	2	1,993	4,490	0,444	2,253
TcCLB.508879.80	hypothetical protein conserved	0	3	2,7	5,0	-1,461	0,000	8,773	0	0	2	1,461	10,234	0,143	7,005
TcCLB.508909.270	hypothetical protein conserved	3	3	19,7	20,0	1,794	2,125	3,986	0	1	2	0,330	2,191	0,151	6,634
TcCLB.508921.30	mitotic centromere-associated kinesin (MCAK), putative	2	3	6,3	7,7	-2,309	2,018	2,231	0	1	2	4,326	4,539	0,953	1,049
TcCLB.508939.50	hypothetical protein, conserved	1	3	3,0	5,3	1,660	1,336	4,832	1	0	2	-0,324	3,172	-0,102	-9,777
TcCLB.508973.30	hypothetical protein conserved	2	3	12,0	14,0	1,047	2,133	3,353	0	1	2	1,086	2,306	0,471	2,124
TcCLB.509001.20	hypothetical protein, conserved	2	3	7,0	7,3	-1,416	1,306	2,394	0	0	2	2,722	3,810	0,714	1,400
TcCLB.509049.51	hypothetical protein, conserved	3	3	18,0	8,7	1,613	8,193	4,918	1	2	3	6,581	3,305	1,991	0,502
TcCLB.509053.100	protein kinase, putative	1	3	6,3	13,7	-1,695	1,091	3,348	1	0	2	2,786	5,043	0,552	1,810
															-
TcCLB.509109.160	hypothetical protein conserved	1	3	4,0	10,3	2,824	2,816	6,909	1	1	3	-0,008	4,085	-0,002	513,997
TcCLB.509119.30	hypothetical protein, conserved	1	3	2,7	6,3	-1,300	7,039	4,626	0	1	3	8,339	5,926	1,407	0,711
TcCLB.509155.30	hypothetical protein, conserved	2	3	19,0	14,0	-2,874	1,848	3,260	0	1	3	4,723	6,134	0,770	1,299
TcCLB.509161.70	Mg transporter, putative	1	3	7,0	8,7	2,925	4,443	5,179	2	1	2	1,519	2,255	0,674	1,485
TcCLB.509167.10	hypothetical protein, conserved	1	3	15,7	18,3	-1,354	1,877	2,514	1	0	2	3,231	3,869	0,835	1,197
TcCLB.509167.200	hypothetical protein conserved	2	3	11,0	57,3	4,815	1,949	8,007	1	1	3	-2,866	3,191	-0,898	-1,113
TcCLB.509167.70	hypothetical protein conserved	0	3	3,7	6,7	-1,591	0,000	4,163	0	0	3	1,591	5,754	0,277	3,617
TcCLB.509215.16	UDP-GlcNAc:PI a1-6 GlcNAc-transferase	1	3	11,3	13,7	-1,476	3,181	2,286	0	1	3	4,657	3,762	1,238	0,808
															908,641
TcCLB.509233.10	hypothetical protein	1	3	9,7	18,7	1,649	1,656	8,116	1	0	2	0,007	6,467	0,001	1
TcCLB.509281.10	hypothetical protein	2	3	8,0	26,3	7,355	1,647	12,560	3	0	3	-5,708	5,206	-1,096	-0,912
TcCLB.509317.10	hypothetical protein, conserved	2	3	30,0	22,0	-2,149	2,309	2,783	0	2	3	4,458	4,931	0,904	1,106
TcCLB.509331.20	hypothetical protein, conserved	3	3	9,7	10,0	-1,523	4,726	5,180	0	3	3	6,250	6,703	0,932	1,072
TcCLB.509641.60	hypothetical protein conserved	2	3	24,3	26,7	1,134	2,585	3,529	0	2	3	1,451	2,395	0,606	1,650
TcCLB.509669.100	hypothetical protein conserved	1	3	4,7	6,0	2,682	1,536	5,530	2	0	2	-1,146	2,848	-0,402	-2,485
TcCLB.509671.20	hypothetical protein conserved	1	3	12,3	17,3	1,716	1,763	5,225	1	0	2	0,047	3,509	0,013	75,063
TcCLB.509693.209	mismatch repair protein, putative	2	3	7,3	6,0	0,000	2,411	2,350	0	1	2	2,411	2,350	1,026	0,975

TcCLB.509693.90	serine/threonine protein kinase, putative	2	3	13,3	9,0	1,145	1,900	3,934	0	0	2	0,756	2,789	0,271	3,692
TcCLB.509695.130	dynein arm light chain, putative	1	3	5,3	5,0	-2,983	1,432	2,611	0	0	2	4,415	5,594	0,789	1,267
TcCLB.509715.26	hypothetical protein, conserved	1	3	7,0	8,0	0,000	5,249	4,460	1	1	2	5,249	4,460	1,177	0,850
TcCLB.509725.139	damage-specific DNA binding protein, putative	1	3	11,3	9,3	2,282	7,321	5,193	1	1	3	5,039	2,911	1,731	0,578
TcCLB.509769.36	palmitoyl acyltransferase 12, putative	2	3	7,7	11,0	2,200	2,600	6,539	1	1	2	0,400	4,339	0,092	10,844
TcCLB.509789.50	rab-like GTPase activating protein, putative	3	3	104,7	105,0	1,751	3,276	3,885	1	2	3	1,525	2,134	0,715	1,399
TcCLB.509805.100	hypothetical protein, conserved	3	3	17,7	13,0	-3,586	3,572	2,306	0	1	2	7,157	5,892	1,215	0,823
TcCLB.509807.20	hypothetical protein, conserved	3	3	16,0	7,3	-1,310	2,505	8,269	0	2	3	3,815	9,579	0,398	2,511
TcCLB.509979.130	hypothetical protein conserved	2	3	3,0	5,0	0,000	1,928	5,012	0	1	2	1,928	5,012	0,385	2,600
TcCLB.510007.50	hypothetical protein, conserved	1	3	16,0	17,3	-2,798	2,863	2,249	0	1	2	5,661	5,047	1,122	0,892
TcCLB.510031.30	hypothetical protein	3	3	14,0	11,0	-1,030	1,385	2,190	1	0	2	2,415	3,221	0,750	1,333
TcCLB.510073.80	hypothetical protein, conserved	2	3	7,3	5,7	-3,032	3,359	3,011	1	2	2	6,392	6,043	1,058	0,945
TcCLB.510091.110	mitochondrial RNA binding complex 1 subunit, putative	1	3	1,3	6,7	0,000	1,212	4,014	0	0	3	1,212	4,014	0,302	3,313
TcCLB.510101.30	60S ribosomal protein L28, putative	1	3	2,7	7,0	-1,892	4,847	4,300	1	1	2	6,740	6,192	1,088	0,919
TcCLB.510105.130	phosphatidylinositol-4-phosphate 5-kinase type II beta, putative	2	3	10,7	11,7	0,000	1,660	2,339	0	1	2	1,660	2,339	0,710	1,409
TcCLB.510121.9	hypothetical protein conserved	1	3	3,0	8,0	1,894	3,817	7,314	1	1	2	1,923	5,420	0,355	2,818
TcCLB.510145.59	exonuclease, putative	2	3	7,7	5,3	1,781	6,606	7,310	0	1	2	4,825	5,529	0,873	1,146
TcCLB.510149.60	hypothetical protein, conserved	3	3	10,7	10,7	-2,160	1,840	5,140	1	1	2	4,000	7,300	0,548	1,825
TcCLB.510181.140	hypothetical protein, conserved	3	3	9,7	5,0	2,766	5,883	5,771	2	3	3	3,117	3,005	1,037	0,964
TcCLB.510275.355	hypothetical protein	2	3	12,7	31,7	1,904	5,029	8,499	1	2	3	3,125	6,596	0,474	2,110
TcCLB.510283.140	UDP-N-acetylglucosamine-dolichyl-phosphate N-acetylglucosaminephosphotransferase, putative	2	3	11,3	20,0	3,925	2,033	6,450	3	1	2	-1,892	2,525	-0,749	-1,334
TcCLB.510285.110	hypothetical protein conserved	0	3	2,7	6,0	0,000	0,000	5,131	0	0	2	0,000	5,131	0,000	0,000
TcCLB.510297.10	hypothetical protein conserved	2	3	14,0	11,3	1,911	1,469	5,503	1	0	2	-0,442	3,593	-0,123	-8,128
TcCLB.510301.40	hypothetical protein, conserved	1	3	4,3	9,7	-1,406	1,212	5,679	0	0	3	2,618	7,085	0,370	2,706
TcCLB.510305.30	lipin, putative	1	3	5,3	11,0	-2,300	2,545	2,558	0	1	2	4,845	4,858	0,997	1,003
TcCLB.510309.30	PTP1-interacting protein, 39 kDa, putative	3	3	15,7	17,3	-1,442	4,051	2,043	0	3	2	5,493	3,485	1,576	0,634
TcCLB.510323.20	hypothetical protein conserved	2	3	8,0	8,3	2,879	2,685	6,527	1	1	2	-0,194	3,648	-0,053	18,758
TcCLB.510329.110	hypothetical protein	3	3	18,0	16,3	-1,756	1,353	2,907	1	0	2	3,109	4,663	0,667	1,500
TcCLB.510351.70	hypothetical protein, conserved	1	3	5,0	7,0	-1,915	1,227	2,152	0	0	2	3,142	4,067	0,773	1,294
TcCLB.510355.269	cytochrome b5, putative	1	3	2,3	6,7	-1,338	1,212	5,552	1	0	2	2,550	6,889	0,370	2,702
TcCLB.510359.320	DNA-directed RNA polymerase III subunit, putative	2	3	23,3	39,3	-2,377	1,497	2,486	0	0	2	3,874	4,863	0,797	1,255

TcCLB.510361.260	mucin-associated surface protein (MASP), putative	3	3	6,7	17,0	6,369	3,249	28,441	2	2	3	-3,120	22,072	-0,141	-7,073
TcCLB.510381.30	hypothetical protein, conserved	3	3	6,3	8,0	-1,150	5,549	4,413	1	1	3	6,698	5,562	1,204	0,830
TcCLB.510491.80	hypothetical protein, conserved	2	3	7,3	7,3	-1,893	1,275	5,577	1	0	2	3,168	7,470	0,424	2,358
TcCLB.510517.90	hypothetical protein conserved	2	3	8,0	7,3	0,000	1,560	5,518	1	0	2	1,560	5,518	0,283	3,537
TcCLB.510531.50	hypothetical protein conserved	2	3	5,0	23,7	1,974	1,560	9,683	1	0	3	-0,413	7,709	-0,054	18,650
TcCLB.510603.100	hypothetical protein conserved	0	3	2,0	6,7	3,219	0,000	5,474	2	0	3	-3,219	2,255	-1,427	-0,701
TcCLB.510629.430	prefoldin subunit, putative	3	3	25,0	17,7	1,340	2,159	3,683	0	2	2	0,819	2,343	0,350	2,860
TcCLB.510659.104	poly(ADP-ribose) glycohydrolase, putative	2	3	31,0	63,3	-1,043	2,922	6,937	1	1	3	3,965	7,980	0,497	2,013
TcCLB.510693.280	mucin-associated surface protein (MASP), putative	2	3	3,7	9,0	1,886	2,551	4,226	1	1	3	0,666	2,340	0,284	3,516
TcCLB.510729.80	hypothetical protein conserved	1	3	10,7	9,0	2,317	2,313	7,054	1	1	3	-0,004	4,737	-0,001	1143,709
TcCLB.510741.50	hypothetical protein, conserved	1	3	2,3	5,0	0,000	5,726	4,056	0	1	2	5,726	4,056	1,412	0,708
TcCLB.510747.60	cytochrome P450 reductase, putative	2	3	12,0	14,7	-3,123	1,845	8,199	0	1	2	4,968	11,323	0,439	2,279
TcCLB.510769.70	hypothetical protein	0	3	6,7	5,0	1,266	0,000	4,326	0	0	2	-1,266	3,060	-0,414	-2,416
TcCLB.511001.130	hypothetical protein conserved	1	3	17,3	21,7	1,211	1,909	3,815	0	0	2	0,698	2,604	0,268	3,732
TcCLB.511029.10	enolase, putative	1	3	6,0	11,0	2,154	2,863	4,328	1	1	3	0,709	2,174	0,326	3,066
TcCLB.511039.39	hypothetical protein conserved	2	3	11,0	5,7	1,843	1,859	5,329	1	1	2	0,015	3,486	0,004	226,186
TcCLB.511131.30	hypothetical protein, conserved	1	3	5,3	18,7	-4,215	1,212	11,400	0	0	2	5,427	15,615	0,348	2,877
TcCLB.511163.24	hypothetical protein conserved	3	3	21,7	21,0	3,482	1,958	5,931	2	1	3	-1,524	2,450	-0,622	-1,608
TcCLB.511191.10	tyrosyl-DNA Phosphodiesterase (Tdp1), putative	3	3	12,3	10,7	4,113	7,982	6,473	2	1	3	3,870	2,360	1,640	0,610
TcCLB.511217.70	hypothetical protein, conserved	1	3	5,0	6,3	-1,778	3,340	11,940	0	1	2	5,118	13,718	0,373	2,680
TcCLB.511277.300	hypothetical protein	2	3	28,7	33,7	-2,197	1,834	2,254	0	1	2	4,031	4,450	0,906	1,104
TcCLB.511279.40	hypothetical protein conserved	2	3	15,3	26,7	1,579	1,422	3,668	0	0	2	-0,157	2,088	-0,075	13,261
TcCLB.511283.80	hypothetical protein, conserved	2	3	20,7	26,7	-1,441	1,174	2,272	0	0	3	2,615	3,713	0,704	1,420
TcCLB.511305.10	5'-3' exonuclease XRN B, putative,exoribonuclease 2, putative	2	3	9,7	8,7	-2,439	1,867	2,979	1	1	2	4,306	5,418	0,795	1,258
TcCLB.511361.30	hypothetical protein conserved	3	3	8,7	6,7	0,000	1,366	4,279	0	0	2	1,366	4,279	0,319	3,132
TcCLB.511409.70	hypothetical protein, conserved	1	3	3,7	5,3	-2,406	1,024	3,283	1	0	2	3,430	5,688	0,603	1,659
TcCLB.511469.20	hypothetical protein conserved	2	3	10,0	6,0	1,565	1,897	3,660	0	1	2	0,332	2,095	0,159	6,307
TcCLB.511501.50	GTPase activating protein, putative	2	3	9,0	5,0	-3,586	3,719	2,310	1	2	2	7,305	5,895	1,239	0,807
TcCLB.511507.100	hypothetical protein conserved	3	3	10,3	11,3	0,000	1,687	9,589	1	1	3	1,687	9,589	0,176	5,683
TcCLB.511531.30	hypothetical protein conserved	2	3	34,0	43,7	1,286	1,942	11,798	0	1	2	0,656	10,512	0,062	16,027
TcCLB.511575.30	hypothetical protein conserved	2	3	9,7	8,3	1,365	1,806	4,312	0	0	3	0,441	2,948	0,150	6,680
TcCLB.511587.75	hypothetical protein	1	3	2,7	7,3	1,062	3,817	6,316	0	1	2	2,756	5,255	0,524	1,907

TcCLB.511589.120	DNA topoisomerase III, putative	1	3	6,7	6,0	0,000	1,312	2,567	0	0	2	1,312	2,567	0,511	1,956
TcCLB.511627.60	vesicle-associated membrane protein (vamp), putative	3	3	6,0	7,0	1,339	2,433	11,417	0	2	2	1,094	10,077	0,109	9,211
TcCLB.511657.79	hypothetical protein conserved	1	3	7,0	11,3	6,849	7,635	9,854	3	1	2	0,786	3,005	0,262	3,824
TcCLB.511717.140	hypothetical protein conserved	2	3	34,7	60,7	2,801	3,011	4,895	3	2	2	0,210	2,094	0,100	9,976
TcCLB.511735.40	hypothetical protein conserved	2	3	8,7	11,0	0,000	1,508	7,687	0	0	2	1,508	7,687	0,196	5,098
TcCLB.511743.10	hypothetical protein, conserved	2	3	9,3	6,3	0,000	4,130	4,117	1	2	2	4,130	4,117	1,003	0,997
TcCLB.511801.34	hypothetical protein, conserved	2	3	78,0	62,3	4,336	6,925	8,789	1	2	2	2,589	4,453	0,581	1,720
TcCLB.511815.60	prefoldin, putative	2	3	8,0	10,0	-1,359	1,486	5,480	1	0	2	2,845	6,840	0,416	2,404
TcCLB.511817.80	protein kinase, putative	3	3	10,7	5,7	0,000	1,523	3,850	0	0	2	1,523	3,850	0,396	2,528
TcCLB.511825.80	I/6 autoantigen, putative	2	3	20,0	27,3	2,278	3,702	11,587	3	1	3	1,424	9,310	0,153	6,537
TcCLB.511857.50	actin-like protein, putative	2	3	6,0	9,7	-1,845	3,035	8,817	0	2	2	4,880	10,662	0,458	2,185
TcCLB.511891.10	trans-sialidase, Group II, putative	3	3	9,0	7,0	2,038	4,997	7,027	1	3	3	2,960	4,989	0,593	1,686